INSTITUTO SUPERIOR DE CIÊNCIAS DA SAÚDE EGAS MONIZ

MESTRADO INTEGRADO EM CIÊNCIAS FARMACÊUTICAS

TÉCNICAS DE EDIÇÃO DE GENOMA COMO ABORDAGEM PROMISSORA NA TERAPIA GÉNICA

Trabalho submetido por
Ana Teresa Gaspar Nabais
para a obtenção do grau de Mestre em Ciências Farmacêuticas

29 de outubro de 2015

INSTITUTO SUPERIOR DE CIÊNCIAS DA SAÚDE EGAS MONIZ

MESTRADO INTEGRADO EM CIÊNCIAS FARMACÊUTICAS

TÉCNICAS DE EDIÇÃO DE GENOMA COMO ABORDAGEM PROMISSORA PARA A TERAPIA GÉNICA

Trabalho submetido por
Ana Teresa Gaspar Nabais
para a obtenção do grau de Mestre em Ciências Farmacêuticas

Trabalho orientado por
Professora Evguenia Bekman

29 de Outubro de 2015

Agradecimentos

Quero agradecer sobretudo à minha família e à minha orientadora Evguenia Bekman toda a dedicação, paciência e força que me deram para alcançar mais esta etapa da minha vida com sucesso.

Obrigada igualmente aos meus amigos por sempre terem acreditado nas minhas capacidades e terem entendido os meus momentos de mau feitio.

Agradeço igualmente a todos os professores do Instituto Superior de Ciências da Saúde Egas Moniz por terem apostado na minha formação profissional.

Não tenho como agradecer todo o apoio que recebi ao longo destes meses de pesquisa e escrita intensiva.

Obrigado

Resumo

As técnicas de edição de genoma consistem na inserção de um ou mais genes manipulados laboratorialmente no interior do genoma celular de um doente, de modo a alcançar a cura de uma doença genética. A sua importância está na capacidade de gerar DSBs e, consequentemente ativar a reparação do DNA lesado com a ajuda de sistemas NHEJ ou HDR. O surgimento da primeira técnica – Zinc-finger nucleases (ZFN) – baseada em fusões de domínios capazes de clivar o DNA das endonucleases de restrição FokI com proteínas ZF, mostrou ao mundo o verdadeiro potencial das técnicas de edição de genoma. Com o passar dos anos foram desenvolvidas mais duas técnicas – TALENs e CRISPR/Cas9 – com o intuito de corrigir e melhorar as desvantagens observadas com o sistema ZFN. TALENs consistem na fusão dos domínios de clivagem FokI e domínios de ligação ao DNA derivados de proteínas TALE. Tal como ZFN, este sistema é capaz de induzir DSBs direcionadas para o DNA que ativam vias de resposta a este dano, permitindo alterações personalizadas do genoma. Por sua vez, o sistema CRISPR/Cas9 derivado do sistema imunitário das bactérias, conta com um crRNA e um tracrRNA específico capazes de invadir DNA invasor. Existem três tipos de CRISPR: I, II e III, sendo o tipo II o mais usado. O aparecimento destas técnicas veio melhorar em muito o tratamento de doenças genéticas e virais tais como o HIV, HBV, HPV, fibrose cística, entre outras. Muitos são os problemas que necessitam de ser resolvidos até se conseguir criar a técnica de edição de genoma ideal para ser aplicada no ser humano.

Palavras – chave: Técnicas de edição de genoma; ZFNs, TALENs, CRISPR/Cas9

Abstract

The genome editing techniques consist in inserting one or more laboratory-manipulated genes within the cell genome of a patient in order to achieve the cure of a genetic disease. Its importance lies in the ability to generate DSBs and thus enable the repair of injured DNA with the help of NHEJ or HDR systems. The appearance of the first technique - Zinc-finger nuclease (ZFN) - based on the fusions of domains capable of cleaving DNA from FokI restriction endonuclease with ZF proteins, showed the world the true potential of the genome editing techniques. Over the years two techniques have been developed - TALENs and CRISPR / Cas9 - in order to correct and improve the disadvantages observed with ZFN system. TALENS consists in the fusion of the cleavage domains of FokI and binding domains derived from protein TALE DNA. As ZFN, this system is able to induce DNA DSBs directed to response pathways that activate this damage, allowing personalised genome changes. In turn, the CRISPR system / Cas9 derived from the immune system of bacteria, and has a specific crRNA and a tracrRNA capable of invading DNA attacker. There are three types of CRISPR: I, II and III, being type II the most used. The appearance of these techniques has improved the treatment of genetic and viral diseases such as HIV, HBV, HPV, cystic fibrosis and others a lot. Many are the problems that must be solved to get the ideal genome editing technique to be applied in humans.

Keywords: genome editing tools; ZFNs, TALENs, CRISPR/Cas9

Índice

Índice de figuras 11

Glossário 13

Introdução 15

Terapia génica 15

Edição de genoma 18

Desenvolvimento 23

ZFNs – Zinc Finger Nucleases 25

TALENs – Transcription Activator-Like Effector Nucleases 33

CRISPR – Clustered regularly interspaced short palindromic repeat 39

Células estaminais e técnicas de edição de genoma 59

HIV e técnicas de edição de genoma 61

Conclusão e perspetivas futuras 65

Bibliografia 69

Índice de figuras

Figura 1 19
Figura 2 25
Figura 3 33
Figura 4 35
Figura 5 40
Figura 6 42
Figura 7 43
Figura 8 45
Figura 9 46
Figura 10 49
Figura 11 50
Figura 12 50
Figura 13 51
Figura 14 53
Figura 15 62

Glossário

Células Jurkat: pertencem à linhagem dos linfócitos T, derivados das células ATLL – células T adultas de leucemia/linfoma- usadas em estudos *in vitro* (Migliorati & Nicoletti, 1994)

Células TIL: linfócitos B infiltrantes de tumores. Normalmente encontrados em tumores, provavelmente refletindo uma resposta imunológica contra os tumores (Anderson, 1990)

Cortes emparelhados: as extremidades obtidas por corte através de enzimas de restrição podem voltar a unir-se por emparelhamento de cadeias complementares simples (Hu et al., 2014)

CRISPR: Clustered regularly interspaced short palindromic repeat (Charpentier, 2015)

Fatores de transcrição OSKM: Os fatores de transcrição denominados OSKM são o Oct4, Sox2, Klf4 e Myc. São proteínas que se ligam ao DNA de células eucarióticas para permitir que haja uma ligação entre a enzima RNA polimerase e o DNA (Carey et al., 2011)

Frameshift: deslocação ou quadro de leitura

Gene CIB1: a proteína que codifica para este gene é a calcium and integrin-binding, pertencente ao grupo de proteínas que codificam para o cálcio. A sua função ainda não foi totalmente determinada, no entanto é conhecida por integrar o DNA dependente de uma cinase e desempenhar um papel de regulação da cinase e da fosfatase (Yuan et al., 2006)

HDR: homology directed repair – reparação homóloga direta (Wu, Ying, Wu, & Capecchi, 2008)

Indels: É um termo utilizado em biologia para descrever a inserção ou deleção de bases no DNA de um organismo (LaFountaine, Fathe, & Smyth, 2015)

NHEJ: non homologous end joing – junção de extremidades não homólogas (Wu et al., 2008)

Nickases: São uma classe de enzimas que reconhecem uma sequência específica na cadeia dupla de DNA, clivando apenas uma das cadeias de modo pré-determinado (Cho et al., 2014)

Ratinhos Knockin: são ratinhos geneticamente modificados, escolhidos devido ao facto de as suas células estaminais embrionárias serem facilmente manipuladas. A técnica

consiste em métodos de engenharia genética que inserem uma proteína codificadora de uma sequência de cDNA dentro do locus de um cromossoma de um organismo (Werner et al., 2006)

Ratinhos Knockout: são ratinhos geneticamente modificados, sendo que um dos seus genes está desativado, substituído ou quebrado por uma sequência de DNA produzida em laboratório. A perda desse gene leva a alterações no fenótipo do ratinho, interferindo na sua aparência, comportamento e outras características físicas e bioquímicas (Kieffer, 1999)

TALEN: Transcription activator-like effector nucleases (Joung & Sander, 2013)

Transfeção: proveniente da junção das palavras transferência + infeção. Termo usado para descrever a produção de vírus infeciosos após a inserção de DNA ou RNA viral em células eucarióticas. Transferência de material genético por meio de um vírus (Charpentier, 2015)

Transgene: é um gene ou material genético transferido entre dois organismos por via natural ou por técnicas de engenharia genética. Descreve um segmento de DNA que contem uma sequência genética isolada a partir de um organismo e introduzido posteriormente num organismo diferente (Dougherty & Parks, 1995)

ZFN: Zinc-finger nucleases (Gupta & Musunuru, 2014)

Introdução

Terapia génica

O conceito de terapia génica surgiu durante os anos 60 e 70, estando desde essa altura em constante crescimento. Em 1972 foi publicado um artigo chamado "Terapia génica para doenças genéticas humanas?" que citou a proposta de Stanfield Roger em 1970 de que o DNA "bom" poderia ser usado para substituir DNA defeituoso em pessoas com doenças genéticas (Mandal, 1990). Nos dias que correm, entende-se por terapia génica a inserção de um ou mais genes manipulados em laboratório, no material genético de células de um doente, de modo a curar uma doença genética (Mandal, n.d.-b).

Os dois tipos de terapia génica existentes são (Mandal, n.d.-b):

- Terapia génica Somática – As alterações de ADN estão confinadas às células somáticas. Assim, a terapia vai provocar efeitos apenas no indivíduo tratado e não na descendência
- Terapia da linha germinal – O gene modificado é incorporado nos gâmetas, causando alterações nos genes herdados pelas gerações futuras. Este tipo de técnica é proibido em vários países devido a problemas éticos

A primeira doente a ser tratada com base na terapia génica foi uma menina com deficiência em adenosina desaminase (ADA), doença que afeta severamente o sistema imunitário e a capacidade de combater infeções. A terapia consistia em retirar os seus glóbulos brancos e neles inserir o gene corrigido de modo a fazer desaparecer a deficiência em adenosina desaminase, e em seguida, reintroduzir as células modificadas no sistema circulatório da doente. Este trabalho serviu de rampa de lançamento para o início de novos estudos e pesquisas sobre este tema. Em 1985 iniciou-se o trabalho de pesquisa para demonstrar como as células de pessoas com esta deficiência poderiam ser modificadas em culturas de tecidos. Os investigadores usaram um retrovírus como vetor para transportar o gene ADA correto para dentro das células (Mandal, 1990).

Atualmente, os principais vetores utilizados para transportar os genes modificados para as células são os vírus. Estes ligam-se a uma célula hospedeira e transferem o seu material genético para o interior da mesma. Aproveitando esta característica peculiar dos vírus, os cientistas começaram a modifica-los, de modo a que no seu interior contivessem

material genético terapêutico em vez de viral. No entanto, é importante que o vírus mantenha a capacidade de infetar as células, dando a oportunidade do DNA benéfico ser inserido nas células hospedeiras (Mandal, n.d.-b). Os vírus utilizados são os retrovírus, os lentivírus, os adenovírus e os adeno-associados. No que diz respeito aos retrovírus, estes são um grupo de vírus cujo RNA é convertido em DNA e integrado no genoma das células que infeta. O RNA viral compreende três genes, *gag*, *pol* e *env*, os quais são flanqueados por elementos denominados repetições terminais longas (LTRs). Estes são necessários para a integração no genoma do hospedeiro, e marcam o início e o fim do genoma viral. As LTRs são igualmente usadas como sequências promotoras, isto é, controlam a expressão dos genes virais. O elemento final do genoma, chamado "sequência de embalamento", permite que o RNA viral venha a ser distinguido de outros RNAs presentes nas células. Uma das limitações deste tipo de vetores é o facto de serem incapazes de infetar células que não se dividem, como as que constituem músculos, cérebro, pulmão e fígado. Assim, quando possível, as células proliferativas do tecido alvo têm de ser removidas, cultivadas *in vitro*, e infetadas com o vetor retroviral recombinante. Seguidamente, as células alvo que produzem a nova proteína são transplantadas de volta para o hospedeiro. Este procedimento foi denominado terapia génica *ex vivo*, e foi inicialmente experimentada em mioblastos e fibroblastos de ratinho (Verma & Somia, 1997). Os lentivírus, pertencentes á classe dos retrovírus, podem infetar tanto as células que sofrem divisão celular, como as que não sofrem. O mais conhecido é o vírus da imunodeficiência humana (HIV), que foi laboratorialmente desativado e desenvolvido como vetor. Tal como os retrovírus, apresenta os três genes *gag*, *pol* e *env*, mas também transporta genes para seis proteínas acessórias, *tat*,*ver*, *vpr*, *vpu*, *nef* e *vif*. Ainda pouco se conhece sobre os possíveis problemas imunitários associados a este tipo de vetor, no entanto, sabe-se que uma injeção de 10^7 unidades infeciosas não induz resposta imunológica (Verma & Somia, 1997). Respeitante aos adenovírus, são uma família de vírus de DNA que infetam tanto células capazes de se dividir como as indivisíveis, podendo causar infeções respiratórias benignas em humanos. Em vez de integrarem o DNA do hospedeiro, são replicados como elementos epissomais no núcleo da célula hospedeira. As células infetadas com adenovírus recombinantes podem expressar o gene terapêutico, mas, devido ao facto de os genes essenciais para a replicação viral serem excluídos, o vetor perde a capacidade de se replicar. Não são muito usados para a terapia génica humana, uma vez que a maior parte da população humana possui anticorpos contra os mesmos. Apesar destes problemas imunológicos, o uso de um grande número de

vetores adenovirais contendo genes específicos, pode provocar a morte de células cancerígenas. Neste caso, a resposta imune contra as proteínas virais irá potenciar a destruição do tumor. Para tentar contornar este problema, foi desenvolvida uma segunda geração de vetores adenovirais, em que os genes adicionados implicados na replicação viral foram excluídos – vetores virais adeno-associados (AAV) (Verma & Somia, 1997). Esta categoria é não patogénica, contendo vírus de DNA de cadeia simples. Compreende dois genes, *cap* e *rep*, que são inseridos entre as repetições terminais que definem o início e o fim do vírus. Os AAV necessitam de genes adicionais para se replicarem, sendo estes geralmente os adenovírus ou o vírus do *herpes simplex*. São capazes de infetar uma enorme variedade de células, sendo que, na presença do gene *rep*, podem atingir as células humanas (Verma & Somia, 1997).

Com o decorrer dos anos, várias foram as tentativas para elevar a fasquia e tentar usar esta técnica para transferir os genes modificados corretamente para a medula óssea, assim como no auxílio para o tratamento do cancro. Os vetores virais criados para transportar marcadores de DNA para as células são capazes de as marcar e permitir aos investigadores um acesso a quais as células TIL mais eficazes, assim como, na confirmação de que o seu uso é seguro e eficaz em humanos (Mandal, 1990).

Com o avanço desta tecnologia foi então permitido o seu uso em recém-nascidos com deficiências genéticas. Os genes são corrigidos e transferidos para células sanguíneas imaturas obtidas a partir dos cordões umbilicais dos mesmos (Mandal, 1990).

As doenças atualmente a serem investigadas para tratamento com uso desta terapia são: fibrose cística, hemofilia, distrofia muscular, cancro e anemia falciforme (Mandal, n.d.-b).

Apesar de tudo, esta técnica tem sido associada a vários problemas ao longo das últimas décadas. Um dos maiores problemas é o facto de não serem conhecidos os seus efeitos adversos a longo prazo e o enorme rol de questões éticas. The Weismann barrier, diz-nos que a terapia génica só pode ser aplicada se for garantida a não transmissão das modificações genéticas aos gâmetas. Esta regra tem de ser obrigatoriamente seguida, uma vez que o seu incumprimento pode levar a alterações hereditárias no genoma, e, consequentemente a transmissão às gerações futuras, em vez de os efeitos serem confinados apenas à pessoa que recebe o tratamento (Mandal, n.d.-a). Outros desafios enfrentados no campo da terapia génica incluem:

- Tempo de terapia curto - Tratamentos múltiplos para alcançar máxima eficácia da terapêutica
- O sistema imunológico pode responder aos vetores virais e perturbar a eficácia da terapia genética
- Os vetores virais podem desencadear nos pacientes reações imunitárias, reações inflamatórias tóxicas ou mesmo haver uma recuperação do próprio vírus, causando doença no paciente
- Risco de induzir crescimento de tumor se o ADN for inserido incorretamente

Estes desafios podem comprometer gravemente todo o processo envolvido nesta terapia, assim como a própria vida do doente (Mandal, n.d.-a, n.d.-b). A terapia génica ainda está em fase experimental, e o seu uso não é, portanto, ainda generalizado (Mandal, n.d.-b).

Edição de genoma

A década passada foi decisiva na rapidez com que a tecnologia de edição de genoma se desenvolveu. Graças ao uso de nucleases modificadas com ligação a domínios específico de DNA de sequência, os investigadores são agora capazes de manipular praticamente qualquer gene de uma gama alargada de vários tipos de células e organismos. Esta rapidez fez com que houvessem meios altamente eficientes, precisos, e agora rentáveis que permitem o estudo da genética funcional e sua terapia, assim como gerar modelos de doenças humanas e animais e aplicações inovadoras na agricultura. (Cox, Platt, & Zhang, 2015; Gupta & Musunuru, 2014; Wu et al., 2008).

Uma edição de genoma eficiente depende de um passo crucial, sem o qual não se atinge uma correta edição de genoma: a criação de uma quebra no DNA no local alvo do genoma que queremos modificar, as conhecidas DSBs – double strand break- (Chen & Gao, 2013; Sander & Joung, 2014). A sua reparação pode ser feita através do método de recombinação homóloga (HDR) ou através da união das extremidades não homóloga (NHEJ), de modo a obter as modificações genéticas desejadas (Sander & Joung, 2014; Sun & Zhao, 2013; Wu et al., 2008). A reparação de uma DSB pode então ocorrer através

de HDR, necessitando de um molde homólogo do DNA (endógeno ou dador) que servirá como guia para a reparação (Sander & Joung, 2014; Sun & Zhao, 2013). No que diz respeito ao DNA reparado por NHEJ, as DSBs vão ser reparadas através da junção das duas extremidades quebradas de DNA alvo, utilizando fatores que voltam a ligar a cadeia sem recorrer a um molde de DNA homólogo (Sander & Joung, 2014). Este processo está muitas vezes sujeito a erros, podendo-se observar a adição de inserções ou mesmo deleções nas DSBs reparadas (Sander & Joung, 2014; Sun & Zhao, 2013; Wu et al., 2008). A sua consequência imediata é a possível rutura da grelha de leitura de uma região codificante ou perda de locais de ligação trans em promotores ou enhancers. Sendo assim, a reparação de uma DSB é muitas vezes imprecisa (Sander & Joung, 2014; Sun & Zhao, 2013; Wu et al., 2008), podendo perturbar a estrutura de leitura translacional de uma sequência de codificação ou mesmo os locais de ligação dos fatores de transcrição aos promotores (Sander & Joung, 2014).

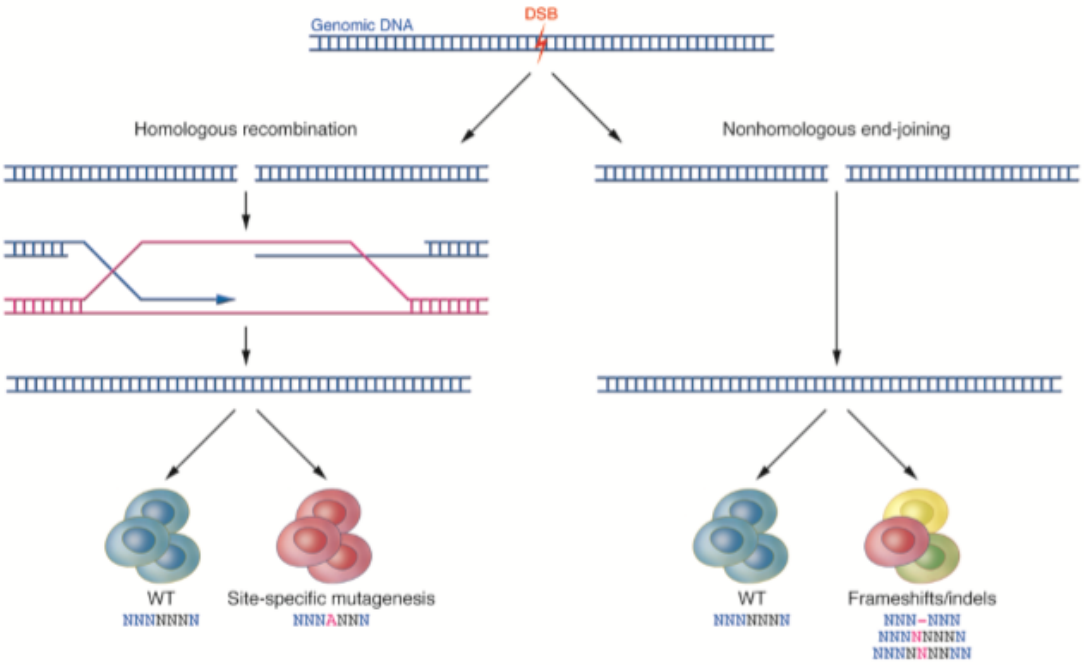

Figura 1: Aquando da criação de DSBs, dois processos de reparação de DNA vão proceder à sua reparação. HDR resulta numa reparação altamente específica utilizando apenas uma fita como molde. Se for necessário, uma sequência de oligonucleótidos exógenos pode ser introduzida para realizar a mutação específica do local. Por sua vez, o método NHEJ produz clones com mutações mais propensas a erros através deste método de reparação (Gaj, Gersbach, & Barbas, 2013)

A recombinação homóloga é o método clássico mais usado para modificação genética. Esta abordagem tem sido amplamente utilizada em células estaminais embrionárias de ratinho para gerar ratos knockout ou knock-in. Uma das suas desvantagens é o facto de apresentar baixa eficácia no que diz respeito a células de mamíferos e de organismos modelo, pois limita drasticamente o seu uso. Isto acontece pelo simples facto de os genomas das células eucarióticas conterem biliões de bases de DNA, dificultando a sua manipulação (Wu et al., 2008). Na sequência da descoberta de que a indução de uma DSB aumenta a frequência de HDR por várias ordens de grandeza, as nucleases alvo têm emergido como o método de escolha para aumentar a eficiência de alterações genéticas HDR-mediadas. As sequências dadoras com < 50pb de homologia, assim como os oligonucleótidos antisense e RNA de interferência têm-se mostrado uma alternativa melhor (Cox et al., 2015). A desvantagem é que normalmente é necessário mais do que um ano para gerar ratinhos geneticamente modificados. Além disso, tentativas similares em usar a recombinação homóloga em células humanas, mostrou que o desafio é muito maior, e que as técnicas alternativas para desativar a expressão dos genes, como oligonucleótidos antisense e os RNA de interferência têm-se mostrado uma alternativa mais adequada; visto que são capazes de induzir mutações, deleções ou mesmo inserções no local alvo. As nucleases específicas de local também permitem a rápida produção de linhas de células e organismos com perda de função. No entanto, estas abordagens podem apenas reduzir transitoriamente a expressão do gene, e o efeito é geralmente mais incompleto e pode afetar genes que não eram o alvo (Cox et al., 2015; Gupta & Musunuru, 2014).

Como já foi referido anteriormente, muitas destas abordagens têm-se mostrado benéficas para desenvolver modelos de doenças humanas, estudar a fisiopatologia de doenças socialmente importantes e modelar uma ampla gama de problemas genéticos. Para isso tem-se recorrido à produção de células estaminais pluripotentes (PCSs) de doentes e a sua diferenciação específica em vários tipos de células somáticas, sendo elas as células estaminais embrionárias (ES) e as pluripotentes induzidas (iPS). Uma das suas vantagens (PCSs) é o facto de serem capazes de se proliferar indefinidamente,

autorrenovar-se e de se transformar em linhagens de células mais diferenciadas (Cox et al., 2015; Gaj et al., 2013).

Uma aplicação particularmente tentadora de nucleases modificadas é o potencial para corrigir diretamente mutações genéticas em tecidos afetados e células para tratar doenças que são refratárias às terapias tradicionais.

Os sistemas de edição de genoma mais importantes hoje em dia são o zinc-finger nucleases (ZNF), transcription activator-like effector nucleases (TALENs) e clustered regularly interspaced short palindromic repeats (CRISPR). Da geração atual de tecnologias de edição de genoma, a que se tem desenvolvido mais rapidamente é a classe das nucleases guiadas pelo RNA, CRISPR (agrupando regularmente espaçamentos entre repetições palindrómicas curtas), que podem facilmente ser direcionadas para praticamente qualquer local do genoma, sendo guiadas por um short guide RNA (gRNA). Estes sistemas abriram portas para a obtenção de uma edição de genoma a nível terapêutico em células e tecidos doentes, o que resulta na remoção ou correção de mutações deletarias ou a inserção de mutações de proteção (Cox et al., 2015; Wu et al., 2008).

Desenvolvimento

Recentemente foram descobertos três novas técnicas de edição de genoma, o que promoveu uma maior esperança e entusiasmo para os melhoramentos da medicina e bioquímica: ZFNs, TALENs e CRISPR/Cas9. Cada uma destas tecnologias utiliza enzimas de restrição para introduzir DSB em locais específicos do genoma humano, através de um guia que pode ser proteínas ou RNA, ligando-se especificamente a uma zona pré destinada do genoma. Estas enzimas oferecem um grande potencial terapêutico, reduzindo significativamente o risco de mutações fora do alvo. Todas elas estimulam os mecanismos endógenos eucarióticos de reparação das DSB, através de NHEJ ou HDR. O seu contínuo desenvolvimento é visto como um avanço significativo em relação aos métodos tradicionais de terapia génica que não possuem tais especificidades. Estas técnicas foram usadas em estudos que visam tratar vários transtornos, incluindo experimentações *in vivo* em mamíferos e até ensaios preliminares em humanos. Para que o potencial terapêutico da edição de genoma seja atingido, é necessário a investigação contínua de novas e melhores estratégias (Khalili, Gordon, Kaminski, Cosentino, & Hu, 2015; LaFountaine et al., 2015).

As promessas iniciais da terapia génica como um tratamento geral para uma grande variedade de doenças diminuiu devido a vários contratempos trágicos. Houve a tentativa de curar com sucesso 17 crianças com imunodeficiência severa- X1 combinada, no entanto, quatro dos pacientes desenvolveram mais tarde sintomas de leucemia. O desenvolvimento dos sintomas de leucemia foram atribuídos aos genes mediados por retrovírus que se inseriram nos locais imprevistos dentro do cromossoma. Em alguns casos, a mutagénese de inserção pode ocorrer numa parte do genoma humano que regula o crescimento da divisão celular, o que pode resultar na proliferação descontrolada de células. Num outro estudo de terapia génica para a deficiência em ornitina transcarbamilase, um paciente morreu devido a uma resposta imunológica contra o vetor do adenovírus usado para entregar o gene de correção. Em ambos incidentes trágicos, o método de entrega terapêutica teve um papel nos efeitos adversos, quer devido a uma resposta imunológica, ou a uma inserção inespecífica (LaFountaine et al., 2015).

Neste trabalho iremos abordar as três técnicas, assim como as perspetivas de melhoramento existentes para o futuro

ZFNs – Zinc Finger Nucleases

ZFNs são sistemas de edição de genoma que utilizam proteínas quiméricas de fusão que compreendem locais específicos nos domínios de ligação de DNA altamente direcionados para "zinc finger", fundindo com um domínio da endonuclease de restrição da enzima FokI proveniente da bactéria *Flavobacterium okeanokoites* (Gupta & Musunuru, 2014; Vasileva, Shuvalov, Garabadgiu, Melino, & Barlev, 2015). As proteínas ZF são comuns em células eucarióticas e estão envolvidas na regulação da transcrição e de interações proteína-proteína. É obrigatório que o DNA seja altamente específico (cada ZF reconhece uma sequência de cerca de 3pb), o que inicialmente gerou interesse nas ZF como ferramentas de edição de genoma. A alteração de um pequeno número de resíduos no interior ou perto deste domínio pode conduzir a alterações na sua especificidade de ligação ao DNA (Joung & Sander, 2013).

Para clivar um local específico no genoma, as ZFNs são concebidas como um par que reconhece duas sequências que atuam no local alvo, um a jusante e outro a montante na cadeia reversa. Após a ligação das ZFNs em ambos os locais, o par dos domínios da FokI dimeriza-se e cliva o DNA nesse mesmo local, originando uma DSB com saliências em 5'. Ao substituir o domínio de reconhecimento natural presente no DNA da FokI com uma sequência de ZF produzida, vai ser formado um ZFN com a capacidade de clivar o DNA num local alvo, baseado na especificidade da ZF (LaFountaine et al., 2015).

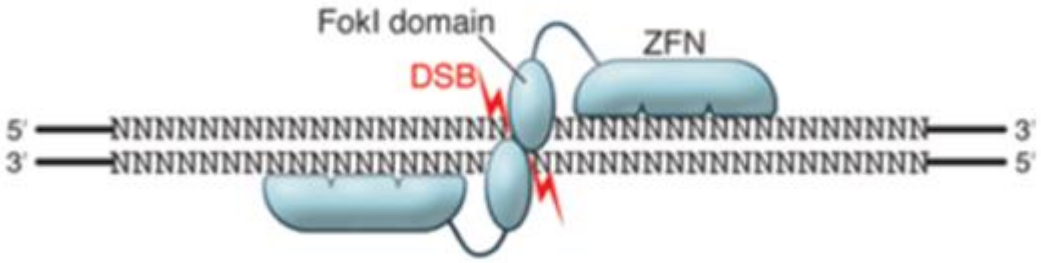

Figura 2: Os comprimentos variáveis dos domínios ZFN ligam-se a sequências do DNA que flanqueiam e posicionam os domínios de FokI de modo a que eles se dimerizem e gerem DSBs entre os locais de ligação (Gupta & Musunuru, 2014)

As DSB são reparadas pelas células usando uma junção de extremidades não homólogas (NHEJ), que pode ocorrer durante qualquer fase do ciclo celular, mas ocasionalmente resulta numa reparação errada ou então, podem usar uma reparação direcionada por homóloga (HDR), que tipicamente ocorre durante a fase S ou G2, quando uma cromatidea-irmã está disponível para servir como um modelo de reparação. A propensão para erros da NHEJ pode ser explorada para introduzir frameshifts (erros na grelha de leitura) na a sequência codificante de um gene, potenciando a sua perda de função por um dos seguintes mecanismos: geração da proteína truncada, não-funcional ou degradação do mRNA não funcional, sendo que este último tem a particularidade de nem sempre ser eficiente. Alternativamente, HDR pode ser utilizada para inserir uma mutação específica, com a introdução de um modelo de reparação contendo a mutação desejada flanqueada por braços homólogos. Em reposta a uma DSB no DNA, HDR utiliza outra sequência de DNA que segue de perto para reparar a rutura. Mecanicamente, HDR pode proceder da mesma maneira que a recombinação homóloga tradicional, usando como modelo de reparação um vetor de DNA de cadeia dupla exógeno (Gupta & Musunuru, 2014).

Cada domínio de ZF reconhece uma sequência de DNA de aproximadamente três a quatro par de bases (pb) e domínios tandem que se podem ligar a uma sequência de nucleótidos mais longa (sendo que normalmente o seu comprimento é múltiplo de 3) que é único dentro do genoma da célula. Assim, uma combinação de seis ZFNs é suficiente para se ligar a 18 pb da sequência original de DNA do genoma alvo, proporcionando especificidade suficiente (Gupta & Musunuru, 2014; Vasileva et al., 2015). É desejável que exista um espaçamento de 16pb entre a sequência de 18pb para que o dímero possa realizar uma correta interação com o DNA. Isto porque estudos anteriores mostraram que com um espaçador de apenas 6pb foram observadas imperfeições relacionadas com a reparação de DSBs dentro e ao redor de um local alvo único de 2x18pb (De Pater, Neuteboom, Pinas, Hooykaas, & Van Der Zaal, 2009).

O domínio da nucleasse pode ser substituído por outros domínios funcionais para manipular os níveis de expressão do gene. O sistema de ZFN foi aplicado com sucesso para a modificação de vários genomas, incluindo plantas, insetos, ratos, ratinhos, porcos e linhas de células humanas e iPSCs (Gupta & Musunuru, 2014; Vasileva et al., 2015)

Relativamente às células estaminais pluripotentes induzidas, Sangamon Biosciences – empresa biofarmacêutica- demonstrou que tanto a tecnologia ZFN, como as iPSCs possuem um enorme valor e versatilidade se forem usadas sinergicamente. Demonstraram igualmente que o aparecimento de mutações levou á reparação do DNA em síndromes de deficiência, tais como Anemia Fanconi, que foram reparadas com o uso de ZFNs e a integração de lentivírus defeituosos para reprogramar hiPSCs – células estaminais pluripotenes humanas induzidas. Estas células reprogramadas mostraram potencial para se formarem células hematopoiéticas progenitoras, que produzem diversas gerações de células corrigidas (LaFountaine et al., 2015). Foi desenvolvido também uma plataforma com um sistema próprio e altamente eficaz para ZFNs e proteínas personalizadas que são modificadas por este método e podem ser facilmente adquiridas (Joung & Sander, 2013).

O uso desta técnica para fins terapêuticos têm sido aplicado principalmente para modificar os genes endógenos numa ampla gama de organismos e células do tipo 2, desordens hereditárias monogénicas, tais como SCID e fibrose cística, bem como infeções adquiridas, como HIV e HBV. Os estudos foram conduzidos numa variedade de linhas celulares humanas e animais e têm incluído mecanismos de interrupção de genes por NHEJ e correção por HDR através do uso de dadores modelo (Joung & Sander, 2013; LaFountaine et al., 2015). Vários tipos de alterações genómicas podem ser introduzidas com esta técnica, incluindo mutações pontuais, deleções, inserções, inversões, duplicações e translocações, proporcionando assim aos investigadores as ferramentas necessárias para realizar manipulações genéticas (Joung & Sander, 2013).

ZFNs estão entre as técnicas de edição de genoma mais desenvolvidas com conclusões de ensaios clínicos de fase precoce para o tratamento de HIV, mostrando ser um forte agente terapêutico, uma vez que demostrou resultados positivos relativamente aos ensaios clínicos realizados em humanos para este problema. (LaFountaine et al., 2015)

O HIV requer um recetor específico, CCR5, encontrado na superfície de células TCD4+, de modo a facilitar a sua penetração viral das células. No entanto, a ausência do recetor CCR5 não resulta em efeitos adversos significativos em seres humanos saudáveis, e tem sido ainda relatado que certas porções da população que não possuem o genótipo CCR5 são resistentes a infeção pelo HIV. Sangamon Biosciences tem enaltecido a capacidade da técnica ZFN conseguir interromper a expressão do gene CCR5 pela

indução de DSBs nos genes. As células TCD4+ mutadas são formadas *ex vivo* e, em seguida, reintroduzidas nos pacientes infetados com HIV, a fim de conferir resistência à replicação viral (Khalili et al., 2015; LaFountaine et al., 2015). ZFN tem a habilidade de perturbar os genes CCR5 e CXCR4, tendo sido testada na infeção por HIV1 que infetava as células TCD4+. Devido ao reduzido número de dadores de CCR5Δ32, tem sido estudada a sua utilização para a obtenção de células sem o gene CCR5 em pacientes HIV1. O seu uso em células humanas é seguro e eficaz (Joung & Sander, 2013; Khalili et al., 2015).

Outras investigações levadas a cabo para esta técnica foram direcionadas para as células que contêm uma mutação no gene IL2R que resulta numa SCID ligada ao cromossoma X. Nestes estudos, dois domínios de ligação ZFN foram desenvolvidos para reconhecer 24 pb que rodeiam o local da mutação e foram transferidas para dentro de K-562 (Leucemia mieloide crónica da medula óssea humana) e para células TCD4+ humanas, juntamente com um plasmídeo dador que realiza uma mutação pontual no fragmento de IL2R. Depois de quatro dias, foi observada uma correção pontual da mutação em 18% das células, confirmando que ocorreu HDR com um plasmídeo dador. A análise mostrou que as células foram transferidas de forma estável e as modificações genéticas foram passadas para as células filhas, demonstrando assim a utilidade clínica de ZFNs para corrigir doenças monogénicas hereditárias. O uso de ZFNs para o tratamento de SCID ligada ao cromossoma X foi igualmente provado com outros estudos, que mostraram altas taxas de adição do gene em vários tipos de células humanas através da entrega de vetores IDLVs -lentiviricos integrasse-defeituosos- (LaFountaine et al., 2015).

O seu uso para corrigir doenças genéticas *in vivo* foi igualmente demonstrado. Como exemplo temos a hemofilia B, sendo esta uma desordem genética que afeta células do fígado causando uma deficiência na produção do fator IX de coagulação do sangue. A maioria das mutações do gene F9 responsável pela produção do fator IX são atribuídos aos exões 2-8 do gene, mostrando assim que qualquer alelo com uma única ZFN pode não ser eficaz. No entanto, investigadores demonstraram que se houver um direcionamento de um intrão a montante das mutações com ZFN, induzidas por HDR, vai conduzir a uma maior expressão da proteína. Esta estratégia foi demonstrada *in vivo* em ratinhos transgénicos hemofílicos, onde o tratamento com ZFNs aumentou as concentrações circulantes de fator IX de coagulação e mostrou um tempo

significativamente reduzido num ensaio de tempo de protrombina parcial ativada quando comparada com mutações em ratinhos não tratados (LaFountaine et al., 2015).

Recentemente foi utilizada uma abordagem alternativa para demonstrar a eficácia *in vivo* de ZFNs noutra doença hereditária ligada ao cromossoma X, a distrofia muscular de Duchenne (DMD). DMD é causada por uma mutação no gene *dystophinI*, que codifica para uma proteína que é essencial para a função muscular. Uma mutação no exão 51 do gene foi apontada em 13% dos pacientes em estudo dos portadores da doença. Numa nova abordagem, os investigadores usaram dois pares de ZFNs que rodeiam cada lado desse exão, a fim de elimina-lo permanentemente após NHEJ e restaurar a expressão da proteína. Esta abordagem foi bem-sucedida em mioblastos de DMD, sendo que foram posteriormente implantados nos membros dos ratinhos para demonstrar uma expressão continua após o transplante (LaFountaine et al., 2015).

A sequência codificadora de um monómero de ZFN é relativamente pequena, aproximadamente 1kb, portanto, eles não são suscetíveis de ser limitados para serem inseridos num vetor (LaFountaine et al., 2015).

Dois estudos subsequentes de ZFNs procuraram identificar potenciais locais fora do alvo para vários pares de ZFN e revelaram vários locais fora do alvo em diversos loci de células tumorais humanas. Assim, os investigadores devem estar cientes da possibilidade de que ZFNs projetados para um propósito particular pode incorrer em eventos fora do alvo. Uma estratégia para reduzir estes eventos é a utilização de um par de ZFNs com domínios FokI distintas e que são obrigatoriamente heterodímeros. Isto impede que um único ZFN se ligue a dois locais adjacentes fora do alvo e que gere uma DSB. Em vez disso, a única maneira de ocorrer DSB fora do alvo, é se ambas as ZFNs de um par se ligarem adjacentemente e, assim permitirem que o dímero FokI se forme. Outra estratégia que foi demonstrada para reduzir eventos fora do alvo, é a introdução de proteínas ZFN purificadas nas células (Gupta & Musunuru, 2014).

A especificidade do DNA alvo verificou ser a maior determinante, tanto na atividade como na toxicidade de ZFNs. Os esforços para melhorar a qualidade desta técnica incluem o desenvolvimento de domínios de clivagem de heterodímeros, otimizando o espaçamento requerido entre duas secções de DNA alvo através de um ligante produzido, e aumentando a atividade de clivagem através de uma evolução direcionada da nuclease FokI. O desenvolvimento de nucleases heterodímeras é talvez a contribuição mais importante para o aumento da sua especificidade. Anteriormente, homodímeros de ZFN (dois monómeros da esquerda ou dois da direita) podiam-se ligar

e clivar o DNA num local específico fora do alvo. No entanto, ZFNs individuais integram uma maior quantidade de DNA e, portanto, conferem uma maior especificidade (LaFountaine et al., 2015).

Como referido anteriormente, os vetores virais de entrega são os mais comuns para servir os requisitos de ZFNs, devido à sua capacidade natural para entrarem nas células e propagar-se no seu genoma. ZFNs devem ser transitórias dentro de uma célula, apenas estando o tempo suficiente para produzir uma DSB. No entanto, os vetores virais não são os mais indicados, uma vez que produzem episssomas, podendo estes vir a desintegrar-se dentro das células. Como foi referido anteriormente, os vetores de lentivírus incorporam diretamente no DNA do hospedeiro, tendo vindo a ser evitados devido a várias tragédias que envolveram a morte de pacientes em ensaios clínicos. Assim, um vetor viral modificado tal como os IDLV – vetor lentiviral de integração deficiente-, vetores adenovirais e virais adeno-associados (AAV), foram fortemente explorados nos últimos anos (LaFountaine et al., 2015).

Dois grupos de investigadores avaliaram a entrega de uma cassete de expressão de ZFNs através de vetores lentivirais de integração deficiente. Em ambos os casos, uma sequência de dador também foi entregue para induzir a correção do gene através de HDR. As correções dos genes foram tão elevadas como 6 e 12% de células em cada estudo, dependendo do tipo de vetor utilizado e células avaliadas, no entanto, integração não intencional foi igualmente observada com IDLVs. Além disso, ambos os estudos foram realizados *in vitro* e assim foi necessário continuar a estudar o facto de IDLV ser ou não um método de entrega adequado para ZFN *in vivo*. IDLVs permanecem uma opção atraente daqui em diante devido a avanços contínuos para reduzir o risco de mutagénese de inserções, a sua capacidade de transferir células divisíveis e não divisíveis, a suba baixa imunogenicidade e imunidade pré-existente. AAV e rAAV – vetores recombinantes adeno-associados- são opções atraentes para a entrega de ZFNs e têm sido estudados por vários investigadores. AAV conseguem transfetar uma grande variedade de células em divisão e não divisão não patogénicas. A principal limitação dos AAVs tradicionais é o seu tamanho relativamente pequeno – aproximadamente 4,7kb. Felizmente, as sequências que expressam ZFNs monoméricos apresentam aproximadamente 1kb de tamanho e assim, com uma conceção cuidadosa, dois monómeros juntamente com um DNA molde opcional de dador podem ser acomodados num único vetor. Integrar todos os componentes num único vetor resulta numa maior especificidade e na redução dos efeitos fora do alvo em comparação com a integração usando vários vetores AAV. Vetores de

rAAV que contêm dois monómeros de ZFN, bem como um DNA molde de dador, foram usados com sucesso em modelos de ratinhos.

Outros estudos relataram a utilização de vetores de adenovírus com capacidade para uma entrega altamente eficiente de ZFNs, no entanto a forte resposta imunitária provocada por estes vetores continua a ser uma preocupação clínica. Além disso, verificou-se que os AAV funcionam de uma forma mais eficiente em alguns tecidos do que noutros (LaFountaine et al., 2015).

Apesar das vantagens da edição de genoma com ZFNs, existem também desvantagens, tais como o facto de não ser simples a montagem dos domínios de ZF de modo a se ligarem a um troço alongado de nucleótidos com alta afinidade. Este problema tornou difícil para os não especialistas conseguirem projetar ZFNs de modo fácil e rotineiro. Para superar estas dificuldades, um grupo de especialistas desenvolveu uma biblioteca aberta com componentes de ZF e protocolos para realizar quadros que identificassem ZFNs que possuam alta afinidade na ligação da sequência desejada. Apesar de facilitar o trabalho, pode levar ainda meses para que os não especialistas sejam capazes de obter ZFNs otimizados. Outra potencial desvantagem é a escolha do local alvo, visto que esta limita os componentes de ZFN de código aberto, podendo estes apenas ser usados para atingir locais de ligação a cada 200pb na sequência aleatória de DNA. É de notar que as fontes comerciais de ZFNs oferecem densidades mais elevadas, com a capacidade de atingir locais de ligação a cada 50pb em sequências de DNA aleatórias. Embora isto possa não ser um problema se um investigador pretender fazer um knockout de um gene, uma vez que um quadro de leitura pode ser introduzido em qualquer lugar no início da sequência de codificação do gene, podendo produzir o resultado desejado, pode apresentar desafios se um local particular é requerido, por exemplo, para obter uma mutação específica. Desde a introdução da biblioteca de código aberto, plataformas alternativas para obter uma engenharia otimizada, cada uma com graus diferentes de velocidade, flexibilidade de escolha do local, e as taxas de sucesso foram aparecendo (Gupta & Musunuru, 2014). Por último, uma preocupação significativa acerca da utilização de proteínas concebidas para introduzir DSB no genoma, é que eles vão fazê-lo não só no local desejado, mas também em locais fora do alvo, podendo causar níveis elevados de toxicidade (De Pater et al., 2009; Gupta & Musunuru, 2014). Os níveis elevados de toxicidade podem ser causados não só por efeitos fora do alvo mas também por níveis de expressão bastante elevados (De Pater et al., 2009). É ainda de salientar que, nos casos em que são observados efeitos nocivos, estes são, pelo menos em parte

consequência da manipulação *in vitro* de células alvo, e não apenas um efeito indesejado induzido somente pelas ZFNs (De Pater et al., 2009)

TALENs – Transcription Activator-Like Effector Nucleases

TALENs é um sistema de edição de genoma aplicado com sucesso na modificação do genoma em plantas, mosca da fruta, peixes, anfíbios, grilos, lombrigas, porcos, ratos, ratinhos, coelhos, células humanas cancerígenas, iPSCs e possivelmente a muitos outros organismos no futuro (Joung & Sander, 2013; Sun & Zhao, 2013; Vasileva et al., 2015). Tem sido utilizada com o intuito de substituir as ZFNs (Joung & Sander, 2013). A componente mais importante deste sistema é uma proteína TALE com ligação a um local específico do DNA obtida através de uma bactéria patogénica de plantas– *Xanthomonas* ou da bactéria *Ralstonia*, podendo ambas ser utilizadas para a edição específica de genoma (Gaj et al., 2013; Joung & Sander, 2013; LaFountaine et al., 2015; Vasileva et al., 2015). Estas TALEs são posteriormente inseridas nas células do organismo hospedeiro através de um sistema de secreção, ligando-se ao DNA genómico e alterando a transcrição das células hospedeiras, facilitando assim a colonização pelas bactérias patogénicas (Joung & Sander, 2013).

Figura 3: Figura 3: Dímero de TALEN ligado ao DNA. Os locais alvo TALEN consistem em dois locais de ligação TALE separadas por uma sequência espaçadora de comprimento variável – 12 a 20pb-. TALEN podem ser projetadas para reconhecer locais alvo únicos (Gaj et al., 2013)

O domínio de ligação de DNA a uma TALE representa um conjunto de repetições de 33-35 aminoácidos (Gaj et al., 2013; Joung & Sander, 2013; Sun & Zhao, 2013; Vasileva et al., 2015). A repetições TALE apresentam sequências similares que diferem apenas em dois aminoácidos – posições 12 e 13- altamente variáveis, e que formam uma matriz para o reconhecimento de nucleótidos específicos e consequentemente para qualquer sequência de DNA (Joung & Sander, 2013; LaFountaine et al., 2015; Vasileva et al., 2015). investigadores observaram que o número de repetições de uma matriz corresponde ao comprimento do seu local alvo, permitindo-lhes deduzir que existe uma relação simples entre os aminoácidos altamente variáveis (12 e 13) e a base de ligação para cada repetição (Joung & Sander, 2013; Sun & Zhao, 2013). A construção destas matrizes de DNA que codificam para as repetições TALEN podem ser desafiadoras devido à exigência de juntar várias sequências de repetição quase idênticas. Diferentes plataformas foram projetadas para facilitar a montagem de plasmídeos que codificam para matrizes de repetição TALEN (Joung & Sander, 2013). Existem quatro conjuntos de aminoácidos que se repetem, denominados RVD-repeat variable di-residue-: Asn-Asn, Asn-Ile, His-Asp, Asn-Gy; sendo suficientes para o reconhecimento específico de um dos quatro pares de base do DNA- guanina (G), adenina (A), citosina (C) e timina (T)-, gerando consequentemente uma TALE com propriedades únicas (Gupta & Musunuru, 2014; LaFountaine et al., 2015; Vasileva et al., 2015).

O segundo elemento desta fusão é o domínio de fusão das nucleases de uma endonuclease de restrição FokI ou outro domínio funcional, capazes de introduzir alterações específicas no genoma (Vasileva et al., 2015). Assim, o código RVD tornou possível criar este novo sistema de engenharia específica de local com base em nucleases que fundem com o domínio de repetições TALE e com um domínio de endonucleases de FokI, denominadas TAL effector nucleases – TALENs (Gupta & Musunuru, 2014; LaFountaine et al., 2015; Mahfouz, Piatek, & Stewart, 2014).

As TALENs são semelhantes ao sistema ZFN no sentido em que também compreendem um domínio de nuclease não específica FokI fundida com um domínio de DNA e podem igualmente gerar uma DSB num local alvo específico do genoma – com aproximadamente 30-40pb-, e assim, ser utilizadas para knockout de genes ou gerar mutações do mesmo modo. Em comparação com ZFNs, TALENs têm sido muito mais fáceis de projetar devido ao uso do código RVD, sendo também muito mais rapidamente construídas, podendo levar apenas dois dias, com um rendimento de centenas de cada vez

(Gupta & Musunuru, 2014; Joung & Sander, 2013; LaFountaine et al., 2015; Sun & Zhao, 2013). A sua fácil projeção, alta taxa de atividade de clivagem e capacidade de efetuar segmentações quase ilimitada tem vindo a estimular ainda mais as investigações para o seu uso em aplicações terapêuticas (Joung & Sander, 2013; LaFountaine et al., 2015; Mahfouz et al., 2014).

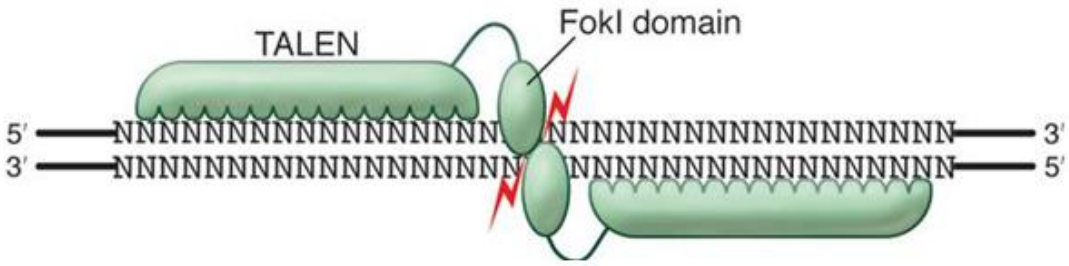

Figura 4: Ligação heterodimérica de TALENs, que como a técnica de ZFNs, existem regiões de ligação de comprimento variável para gerar DSBs entre os locais de ligação (Gupta & Musunuru, 2014)

Embora as técnicas de ZFN e TALEN não tenham sido diretamente comparadas, estudos demonstraram que ambas são capazes de clivar o DNA com eficiência semelhante quando orientadas para o mesmo gene, assim como introduzir eficientemente alterações orientadas em vários organismos modelo, que antes eram difíceis ou impossíveis de manipular geneticamente, como ratos ou mesmo porcos. Além disso, as TALENs também têm sido usadas para modificar genes endógenos de vaca, grilo e bicho-da-seda. A maioria dos estudos realizados utilizaram uma única TALEN para gerar mutações induzidas por NHEJ, no entanto, dois relatórios descreveram o uso de dois pares de TALEN, dirigidos para o mesmo cromossoma, gerando deleções/inserções de grandes segmentos cromossómicos (Joung & Sander, 2013).

Recentemente, uma nova variante do sistema TALE foi desenvolvido, um sistema optogénico LITE – light-inducible transcriptional effectors. Este sistema é composto por dois componentes, sendo que o primeiro é o domínio TALE de *Xanthomonas* com a proteína fotossensível Cry2 (TALE-Cry2) de *Arabidosis thaliana*; e o segundo componente compreende CIB1 (interação em parceria com Cry2), fundindo com um domínio de ligação de um ativador de transcrição – um ativador viral VP64 por exemplo (Vasileva et al., 2015). O tratamento das células com base neste método provoca uma

alteração conformacional na Cry2, facilitando o recrutamento de CIB1: complexo TAD para induzir a transcrição a partir do promotor alvo. Além da regulação da atividade de transcrição, tal sistema pode também ser usado para o direcionamento de modificações epigenéticas específicas de cromatina para locus específicos do genoma. Esta abordagem permite estudar o efeito das modificações da cromatina selecionados na expressão de genes específicos (Vasileva et al., 2015).

Como vimos para ZFNs, o potencial terapêutico das TALENs foi demonstrada em modelos celulares animais para uma grande variedade de doenças genéticas adquiridas, assim como alguns exemplos de perturbações e correções genéticas por NHEJ e HDR (LaFountaine et al., 2015; Sun & Zhao, 2013).

A maioria dos estudos publicados concentram-se em tratar doenças monogénicas herdadas incluindo SCID, DMD, A1ATD, PV, RDEB, anemia falciforme e hemofilia A. A tecnologia TALEN tem progredido de tal modo que os investigadores conseguiram corrigir a mutação pontual mais comum que é responsável pela fibrose cística com o auxílio de hiPSCs (LaFountaine et al., 2015).

Para doenças como a DMD e a hemofilia A, foi mostrado que TALENs podem ser usadas para corrigir uma mutação na ausência de DNA dador para reparação. Isto representa um caso interessante pois a maioria dos casos de hemofilia A são caraterizados por uma inversão cromossómica no gene F8. Para combater tal facto, os investigadores desenvolveram um par de dímeros TALEN que teve como alvo duas extremidades de uma sequência cromossómica de 140 kpb no modelo hiPSCs de hemofilia A (criado pelo mesmo par de TALENs). Após a transfecção com os plasmídeos TALEN, foi observada uma reversão em 1.3% das células. Além disso, quando as iPSCs foram diferenciadas em células endoteliais, também foi observada a expressão de proteínas F8 necessárias para a coagulação do sangue. Estudos recentes sugerem que as tecnologias de edição de genes tais como TALENs podem ser usadas para corrigir erros de genoma mais complicados, tais como inversões cromossómicas de sequência longa (LaFountaine et al., 2015).

Em outros modelos de doenças, a correção dos genes por HDR foi efetuada por inclusão de um DNA dador juntamente com plasmídeos que codificam para as TALEN. A fim de corrigir a mutação no gene da globina causadora de talassemia, os investigadores introduziram um vetor dador TALEN nas hiPSCs por eletroporação. Esta estratégia resultou em eficiências muito elevadas, de 68% e 40% em hiPSCs de dois doentes diferentes respetivamente. As hiPSCs permaneceram pluripotentes e quando diferenciadas foram capazes de expressar a proteína normal (LaFountaine et al., 2015).

TALENs também podem ser eficazes na correção de doenças resultantes de mutações no DNA mitocondrial (mtDNA). As mitocôndrias são organitos que possuem o seu próprio DNA circular contendo 37 genes. A acumulação de mutações no mtDNA está fortemente associado com o processo de envelhecimento e também pode ser transferida para as células germinativas. As mutações nestes genes podem causar distúrbios tais como a miopatia mitocondrial, o síndrome de MERRF, e síndrome de Leigh. Os investigadores projetaram TALENs para atingir especificamente mtDNA, incluindo sequências de localização mitocondrial, em vez de sequências de localização nuclear (LaFountaine et al., 2015).

Tal como nas ZFNs, foram igualmente desenvolvidas técnicas para combater doenças adquiridas com o auxílio das TALENs. O silenciamento do gene CCR5, necessário para a replicação do HIV foi realizada com sucesso em hiPSCs. De igual modo, uma outra proteína celular necessária para a integração do HIV, LEDGF/p75 foi eliminada em células Jurkat onde de seguida, foi confirmada a inibição do HIV (LaFountaine et al., 2015). Vários estudos evidenciaram que o local alvo do gene CCR5 é o mesmo quer para ZFNs quer para TALENs, no entanto, TALENs produzem menos mutações fora do alvo e evidenciam menor toxicidade, ou seja, são uma melhor alternativa (Gupta & Musunuru, 2014).

TALENs foram também desenvolvidas para atingir o DNA viral do HBV e do HPV, de modo a reduzir a infeção induzida por estes. Plasmídeos de TALEN alvo para o HPV foram complexados com uma mistura de polímeros catiónicos e aplicados diretamente no colo do útero de ratinhos transgénicos que apresentava esta infeção e cancro cervical. A terapia reduziu com sucesso as cargas virais e tamanho do tumor. Não foram detetadas mutações fora do alvo e o tratamento foi bem tolerado, sem sinais de resposta inflamatória nem um aumento no número de plasmídeos (LaFountaine et al., 2015).

Relativamente às suas vantagens, as TALENs apresentam reduzida toxicidade comparativamente com as ZFNs, maior facilidade de obtenção e maior flexibilidade devido ao seu domínio de ligação ao DNA (LaFountaine et al., 2015). Este último ponto é explicado pelo facto da matriz de repetições poder ser facilmente estendido para qualquer comprimento desejado (Gaj et al., 2013; Sun & Zhao, 2013). No entanto, as repetições podem também ser vistas como um problema, uma vez que podem-se criar sequências extensas de repetições, sendo um desafio devido ao facto de se poderem

desenvolver repetições idênticas. Para superar este problema, vários métodos foram desenvolvidos para permitir a montagem de repetições personalizadas (Gaj et al., 2013). Na técnica de ZFN existe tipicamente uma ligação de sequências entre 9 e 18pb, enquanto que nas TALENs podemos ligar sequências de 18 ou mais pares de bases. Este facto pode ser simultaneamente apresentado como uma vantagem ou uma desvantagem, uma vez que investigadores afirmam que quanto maior o seu tamanho, menor a sua especificidade (Gupta & Musunuru, 2014). Este ponto pode ser explicado pelo facto de cada monómero TALEN requerer uma cassete de aproximadamente 3kb, o que pode vir a ser um problema no que diz respeito à sua entrega por vetores AAV, uma vez que são muito pequenos- cerca de 4.8kb- para acomodar um par te TALENs monoméricos, tendo estas mais de 6kb (Gupta & Musunuru, 2014). Tendo em conta este aspeto, podemos concluir que os adenovirais são os mais adequados, pois são capazes de transferir de modo eficiente uma TALEN funcional e intata para as células, de modo a que as proteínas resultantes sejam altamente eficientes na produção de DSBs no local alvo desejado (LaFountaine et al., 2015). Além disso, as TALENs são proteínas grandes, apresentando várias sequências repetitivas. Estas sequências são conhecidas por apresentarem instabilidade durante a replicação, reparação e recombinação, o que limita os vetores que são capazes de transferir com sucesso TALENs intactas para um gene de células humanas (LaFountaine et al., 2015). A sua capacidade de ser compactadas e empacotadas nalguns vetores virais vê-se então comprometida, embora este problema possa ser superado através da diversificação das sequências de codificação (Gupta & Musunuru, 2014). Outra vantagem em relação às ZFNs é o facto de haver menos restrições sobre a escolha do local de ação, uma vez que estão disponíveis múltiplos pares TALENs para cada par de base de uma sequência aleatória de DNA (Gupta & Musunuru, 2014).

Tal como referido na técnica de ZFN, os efeitos fora do alvo são igualmente uma forte preocupação para a técnica TALEN, podendo acontecer devido à reparação utilizando a técnica NHEJ, levando até à possível inativação da função do gene. Alternativamente, se for fornecido um de DNA dador de cadeia dupla, pode ser despoletado o mecanismo de HDR, e assim cria-se a oportunidade de introduzir substituições nucleótidas exatas ou inserções no local de quebra ou perto do mesmo (Gupta & Musunuru, 2014; Joung & Sander, 2013).

CRISPR – Clustered regularly interspaced short palindromic repeat

O sistema CRISPR/Cas9, também conhecido como CRISPR/Cas tipo II (Charpentier, 2015), é a mais recente técnica de edição de genoma (Gaj et al., 2013; LaFountaine et al., 2015; Sander & Joung, 2014), bastante promissora nas áreas da biotecnologia e medicina (Charpentier, 2015; Sander & Joung, 2014). É baseada no sistema imunitário das bactérias que as ajuda a proteger de fagos ou plasmídeos através da transferência de genes invasores (Charpentier, 2015; Gaj et al., 2013; Kennedy et al., 2015; Mahfouz et al., 2014; Nihongaki, Yamamoto, Kawano, Suzuki, & Sato, 2015; Vasileva et al., 2015). As bactérias foram as primeiras formas de vida a aparecer na terra e têm revolucionado bastante a biologia molecular visto serem uma fonte quase ilimitada de enzimas, podendo usá-las para diversas aplicações como fabrico de produtos lácteos e produção de substâncias biológicas como vacinas e antibióticos por exemplo (Charpentier, 2015).

As suas capacidades de adaptação às mudanças e condições ambientais, comunicação celular, metabolismo e patogenicidade têm despoletado bastante interesse na área científica, obrigando a uma compreensão mais profunda dos seus mecanismos, a fim de os conseguir reproduzir para aplicações biotecnológicas inovadoras, tais como a edição de genoma e estratégias anti-infeciosas (Charpentier, 2015).

Foi observado que as bactérias eram capazes de se defender de organismos invasores através da cópia de um segmento de aproximadamente vinte pares de bases do DNA invasor para um local do genoma do hospedeiro através de repetições palindrómicas curtas espaçadas entre si – CRISPR (Charpentier, 2015; LaFountaine et al., 2015; Vasileva et al., 2015).

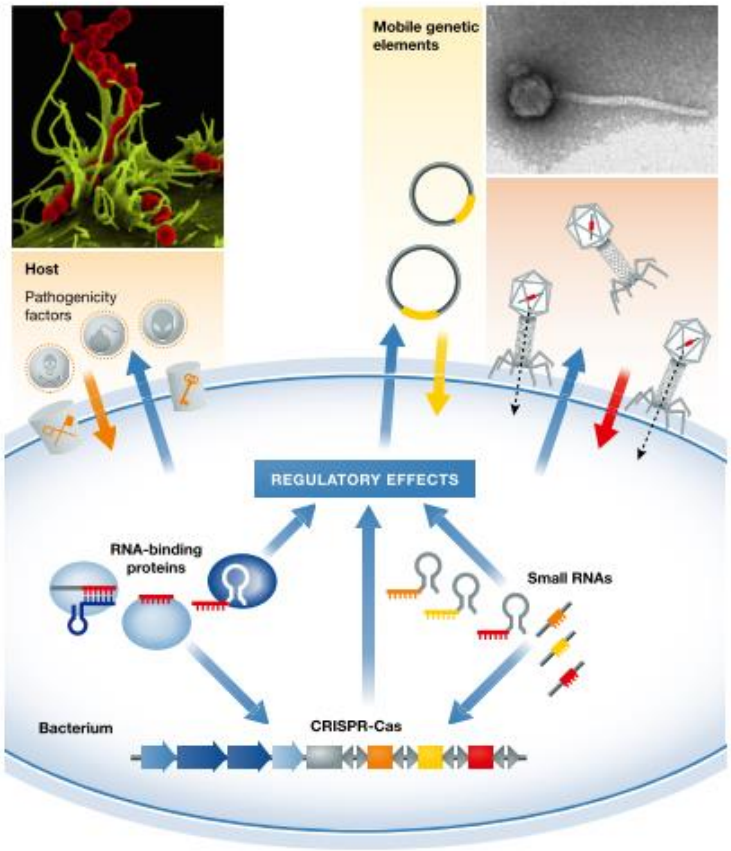

Figura 5: O sistema CRISPR/Cas9 é um sistema de defesa adaptativo mediado por RNA contra elementos genéticos móveis, tais como fagos ou plasmídeos. No caso de Streptococcus pyogenes, o sistema CRISPR/Cas9 tipo II influencia o potencial de virulência do agente patogénico humano, limitando a aquisição de genes de virulência em fagos transportados

Esta transferência tem como objetivo funcionar como uma memória genómica dos patogénios invasores (LaFountaine et al., 2015) e funcionar como um mecanismo de defesa para ataques futuros pois o DNA armazenado nos CRISPR é transcrito em crRNA correspondente (LaFountaine et al., 2015; Vasileva et al., 2015). Este crRNA irá posteriormente reconhecer e ligar-se ao DNA invasor através de emparelhamento de bases, associado a um conjunto de endonucleases endógenas (CAS), gerando uma DSB no DNA invasor e inibindo a sua replicação e integração no DNA (LaFountaine et al., 2015; Sander & Joung, 2014; Vasileva et al., 2015). Compreendemos então que existem dois possíveis cenários de interação com os invasores: a memorização de um elemento genético após uma primeira infeção ou a destruição do mesmo elemento após uma segunda infeção. No primeiro caso, o elemento é reconhecido pelas proteínas Cas que

inserem uma sequência curta do DNA invasor numa matriz de CRISPR localizada no genoma do hospedeiro (Charpentier, 2015). Estas sequências que são integradas constituem então uma memória genética que impede que o hospedeiro seja reinfectado. Vai haver então uma memorização das sequências invasoras por parte do hospedeiro, transcrevendo-as numa molécula de RNA que é posteriormente submetida a um ou dois eventos de processamento que darão origem a uma cadeia curta de crRNA, contendo uma parte do elemento genético memorizado (Charpentier, 2015; Nihongaki et al., 2015).

Logo após os investigadores se terem apercebido deste fenómeno começou-se a avaliar este sistema para células eucarióticas. A endonuclease Cas9 foi otimizada para incluir sinais de localização em células humanas, sendo capazes de se ligar a qualquer sequência alvo de DNA de vinte pares de bases (LaFountaine et al., 2015; Vasileva et al., 2015).

O sistema CRISPR/Cas9 é então composto por uma endonuclease Cas9 e duas moléculas de RNA curtas – crRNA e tracrRNA- (Gupta & Musunuru, 2014; Mahfouz et al., 2014; Robert, Barbeau, Éthier, Dostie, & Pelletier, 2015; Sally, n.d.; Sander & Joung, 2014; Vasileva et al., 2015), sendo que o crRNA é o responsável pela orientação da endonuclease Cas9 para o DNA alvo (Mahfouz et al., 2014; Sander & Joung, 2014). Estas moléculas de RNA pode ser combinadas com uma molécula de RNA guia – gRNA- com uma sequência de vinte nucleótidos (Ran et al., 2013; Sander & Joung, 2014; Vasileva et al., 2015) capaz de dirigir especificamente a proteína Cas9 para o seu destino (Mahfouz et al., 2014; Sander & Joung, 2014). O crRNA é obtido a partir da transcrição de pequenos segmentos de DNA exogeno, interligados entre si dentro do loci genómicos de CRISPR (Gaj et al., 2013). Os crRNA vão emparelhar com os tracrRNA específicos da sequência, orientando as Cas9 para certos "protospacers" localizados no DNA alvo, conseguindo assim o seu silenciamento (Gaj et al., 2013; Mahfouz et al., 2014; Sander & Joung, 2014). Uma alternativa para conseguir o seu silenciamento é associar as proteínas Cas9 com complexos de proteínas ribonucleicas de modo a interferirem com os ácidos ribonucleicos do DNA invasor, conseguindo assim o seu silenciamento (Charpentier, 2015)

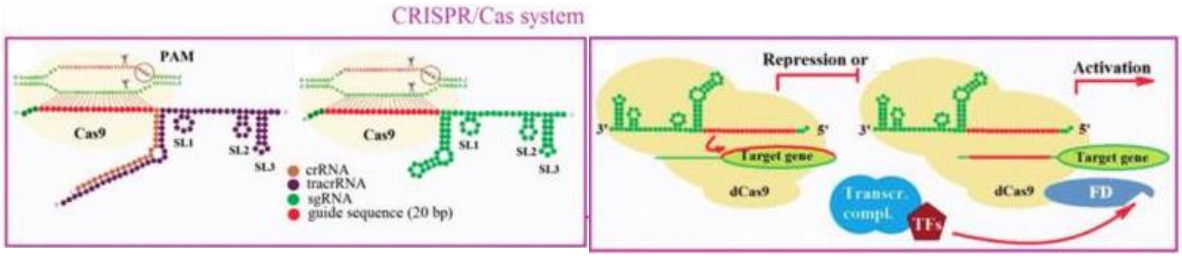

Figura 6:Representação esquemática do sistema CRISPR/Cas9, que consiste num domínio de uma nuclease Cas9 que se agrega a um crRNA e tracrRNA, de modo a que esta seja direcionada para o local alvo. Na imagem, o local alvo é indicado por uma tesoura. Podemos observar o PAM dentro dos círculos. Vai haver uma ligação de dCas9 específica do local a um sgRNA, podendo inibir a interação com os fatores de transcrição, causando uma repressão do gene da região promotora. Após a fusão do sgRNA com o fator de transcrição a transcrição vai ser facilitada (Vasileva et al., 2015)

O tracrRNA é um componente essencial do complexo ribonucleico do sistema CRISPR/Cas9, atuando como uma ativador necessário para processar todo o crRNA e também como um componente essencial e indispensável do complexo de DNA. Tem-se mostrado igualmente um fator importante na função reguladora da expressão em genes endógenos (Charpentier, 2015).

O crRNA pode ser usado para qualquer um dos tipos do sistema CRISPR/Cas9, enquanto que o tracrRNA apenas pode ser usado para o tipo II (Charpentier, 2015).

O reconhecimento da proteína Cas9 requer uma sequência do crRNA e de um PAM – motivo adjacente ao protospacer – localizado a jusante da região do crRNA (Gaj et al., 2013; Gupta & Musunuru, 2014; Mahfouz et al., 2014; Nihongaki et al., 2015; Vasileva et al., 2015). O sistema CRISPR/Cas9 pode então ser redirecionado e assim conseguir clivar praticamente qualquer sequência de DNA após a programação do crRNA para o nosso alvo (Gaj et al., 2013).

A escolha da plataforma de gRNA é importante, pois a eficácia da atividade da Cas9 para qualquer locus pode ser influenciada pela sua estrutura (Sander & Joung, 2014).

A maioria dos estudos têm relatado o uso de um único gRNA, sendo este a fusão do crRNA programado e de uma parte de um tracrRNA. No entanto, é também relatado o seu uso em separado, mostrando taxas de edição substancialmente mais baixas do que as utilizadas com um único gRNA, sendo então este sistema mais ativo do que os sistemas com gRNA duplos (Sander & Joung, 2014). Além disso, os gRNAs individuais conseguem albergar comprimentos variáveis da sequência tracrRNA nas suas extremidades 3’ (Sander & Joung, 2014).

Outro passo importante neste sistema de edição de genoma é a escolha dos locais alvo e do promotor utilizado (Sander & Joung, 2014). O promotor que expressa o gRNA pode limitar as opções para potenciais locais alvo do DNA (Sander & Joung, 2014). Vejamos um exemplo – o promotor U6 da RNA polimerase III requer uma guanina (G) e o promotor T7 requer duas na mesma extremidade (Gupta & Musunuru, 2014; Sander & Joung, 2014; Vasileva et al., 2015), como podemos observar na figura abaixo.

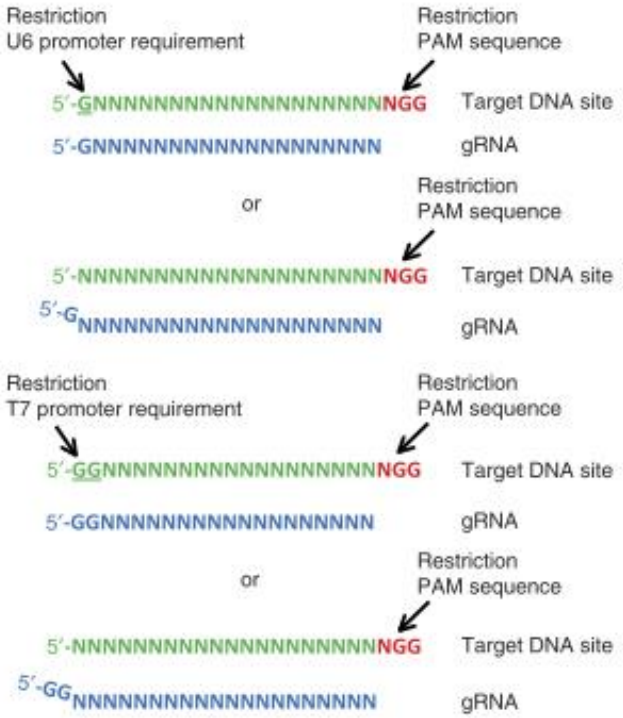

Figura 7: Intervalos de potenciais locais alvo para a expressão dos gRNAs apartir de um promotor U6. Restrições nas sequências de DNA alvo impostas pela exigência de uma guanina (G) no primeiro nucleótido 5' dos gRNA – letras azuis – voltado para o local de DNA e pela necessidade de um NGG – sequência PAM a vermelho – adjacentes à região de complementaridade das sequências no local alvo – letras verdes - .Relativamente à figura de baixo, podemos observar intervalos nos locais alvo potenciais para expressar gRNAs a partir de um promotor T7. Observam-se restrições nas sequências de DNA alvo impostas pela exigência de um GG com os dois primeiros nucleótidos do gRNA direcionado para o local alvo do DNA e pela necessidade de um NGG adjacente ao local alvo na região complementar (Sander & Joung, 2014).

Foi então necessário recorrer a estratégias que reduzissem as restrições para a escolha do local alvo, fazendo gRNAs incompatíveis para estas posições, escolhendo locais sem ter em conta as identidades das primeiras duas bases na extremidade 5'(Sander & Joung, 2014). Outra estratégia bastante utilizada é anexar a guanina, ou as duas guaninas, necessárias á extremidade 5' do gRNA, codificando assim transcrições de

gRNA que são um ou dois pares de bases mais longos, como podemos observar no final de cada cadeia da imagem acima (Sander & Joung, 2014). Ambas as estratégias têm sido usadas com sucesso para produzir gRNAs ativos de eficiência variável em atividades de genoma de edição com Cas9 (Sander & Joung, 2014).

A primeira Cas a ser encontrada e demonstrada laboratorialmente como sendo eficiente foi a proveniente de *Streptococcus thermophilus* (Charpentier, 2015), na qual se observou laboratorialmente o reconhecimento e clivagem de DNA de plasmídeos invasores através da interferência com e sem PAM (Charpentier, 2015; Nihongaki et al., 2015). Em relação ao agente patogénico *Francisella novicida*, observou-se que a expressão dos fatores de virulência era regulada através de tracrRNA, orientando um complexo formado por um scaRNA e uma proteína Cas9. Este complexo mostrou-se um meio alternativo capaz de proporcionar um silenciamento específico do RNA mitocondrial em células e organismos (Charpentier, 2015).

O uso de cortes emparelhados é outra alternativa testada em ratinhos pelos investigadores para tentar reduzir as mutações fora do alvo. O seu uso pode reduzir este tipo de atividade em cerca de 50 a 1500 vezes em linhas celulares, facilitando o knockout de genes sem colocar em causa a eficiência da clivagem do alvo (Ran et al., 2013).

O sistema CRISPR/Cas9 está classificado hoje em dia em três tipos principais – tipo I, II e III, e onze subclasses, sendo que o tipo II-A é o mais conhecido e estudado. Ambos os tipos I e III podem ser encontrados em bactérias e no reino *Archaea*, ao passo que o tipo II é apenas encontrado em bactérias (Charpentier, 2015). Todos os três têm em comum o facto de serem baseados na imunidade adaptativa como foi referido no início do capítulo. O que os diferencia é a proteína Cas utilizada e o seu mecanismo de ação para desempenhar a memorização, obtenção do crRNA e possíveis interferências com o DNA invasor (Charpentier, 2015). Os sistemas I e III são bastante semelhantes, sendo que ambas as suas interferências envolvem crRNP – complexos de proteínas ribonucleicas – sendo que cada um é composto por um crRNA guia e um complexo de proteínas Cas. Comparativamente, o tipo II é específico para um crRNA duplo com um RNA adicional – tracrRNA – formando um complexo de DNA que interfere com uma única proteína Cas9 (Charpentier, 2015).

O sistema CRISPR tipo II de *S.pyogenes* foi adaptado para induzir DSBs específicas de sequência e edição de genoma segmentado. Na sua forma mais simples, dois componentes devem ser introduzidos e expressos em células onde se pretende

realizar a tão desejada edição de genoma: a nucleasse Cas9 e um RNA guia – gRNA, resultado da fusão de um crRNA e um tracrRNA (Sander & Joung, 2014).

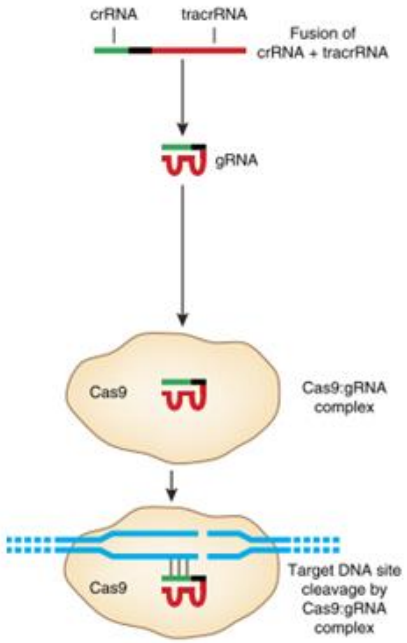

Figura 8: O sistema CRISPR/Cas9 é a técnica de edição de genoma normalmente utilizada para fundir um crRNA e parte da sequência de um tracrRNA. Os complexos gRNA:Cas9 vão mediar a clivagem dos locais alvo do DNA complementares à extremidade 5' de 20 nucleótidos do gRNA e que se encontram lado a lado com uma sequência PAM (Sander & Joung, 2014)

Os vinte nucleótidos no final da cadeia 5' do gRNA correspondem à porção do protospacer do crRNA, e servem para direccionar a Cas9 para um local específico de DNA alvo baseando-se no emparelhamento RNA-DNA por complementaridade. Estes locais alvo devem situar-se imediatamente antes de uma sequência 5'NGG que corresponde ao PAM. Assim, com a utilização destes sistemas, a nuclease Cas9 pode ser direcionada para qualquer sequência de DNA na forma de N_{20}-NGG pelo simples facto de alterar os primeiros vinte nucleótidos do gRNA correspondentes à sequência de DNA alvo (Sander & Joung, 2014).

Foram avaliados outros sistemas tipo II noutras bactérias, uma vez que também são capazes de realizar a edição de genoma segmentado. No entanto, o sistema do *S.pyogenes* é o mais conhecido e utilizado. Esta plataforma baseada na Cas9 funciona também de um modo eficiente para diferentes tipos de células e organismos. Foi

demonstrada a sua funcionalidade em genes endógenos de bactérias, linhas celulares de cancro humano, em células estaminais pluripotentes humanas, bem como em organismos como o peixe-zebra. Foi igualmente utilizado para modificar genes de arroz, tabaco, leveduras, sapos, entre outros. Este sistema é capaz de introduzir DSBs em vários locais em paralelo, utilizando a Cas9 com múltiplos gRNAs, apresentando uma vantagem em relação às ZFN e TALEN (Sander & Joung, 2014).

A simplicidade com que a Cas9 pode clivar o DNA também inspirou para a geração de grandes bibliotecas de gRNAs utilizando uma síntese de oligonucleótidos baseada em matrizes. Estas bibliotecas podem ser modificadas para abranger múltiplos gRNA para quase todos os genes de um organismo hospedeiro, facilitando assim a seleção genética. Cas9 apresenta várias variantes que cortam apenas uma das cadeias do DNA alvo (conhecidas como nickases), sendo estas úteis para a edição de genoma. A introdução de uma mutação no gene D10A ou H840A no RuvC1 por exemplo, pode resultar na geração de nickases que cortam tanto as cadeias alvo de DNA complementares como as não complementares, como podemos observar nas imagens abaixo indicadas.

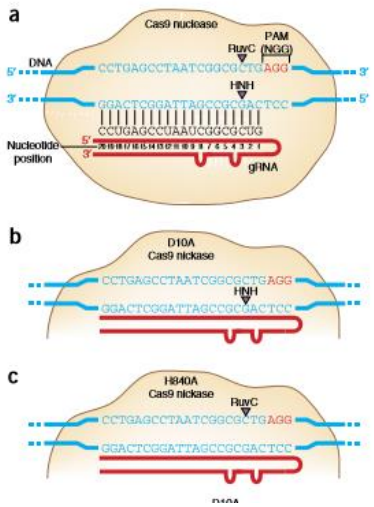

Figura 9: (a) A nuclease Cas9 cria DSBs em locais alvo do DNA. (b) Seguidamente uma nickase será criada pela mutação de um domínio da nuclease RuvC e irá clivar apenas a cadeia de DNA que é complementar e reconhecida pelo gRNA. (c) Temos a criação de uma nickase criada pela mutação do domínio da nuclease HNH. Esta nickase vai clivar apenas a cadeia de DNA que não interage com o gRNA (Sander & Joung, 2014)

Um dos pontos importantes que temos de ter em consideração é o facto de estas nucleases serem capazes de produzir efeitos fora do alvo, ponto que irei focar mais adiante quando falar sobre as desvantagens da técnica. Para avaliar a especificidade desta técnica, têm sido criados várias variantes de gRNA contendo um a quatro desemparelhamentos de nucleótidos na região de complementaridade e avaliada a sua capacidade de dirigir Cas9 para genes alvo de células humanas. O efeito dos desemparelhamentos nem sempre é previsível com base na sua localização dentro da região alvo do gRNA, sendo que por regra, alguns erros no extremo 5' podem ter efeitos indesejados, enquanto que alguns na extremidade 3' não afetam significativamente a especificidade da Cas9. Temos de ter em consideração que alguns destes desemparelhamentos são melhor tolerados no final da região 5' dos vinte nucleótidos da sequência de gRNA do que na extremidade 3' (Sander & Joung, 2014). Além disso, nem todas as substituições de nucleótidos podem ter efeitos indesejados (Sander & Joung, 2014).

As vantagens do sistema CRISPR/Cas9 passam por uma alta versatilidade, eficácia, especificidade e facilidade de uso, em comparação com as outras duas técnicas abordadas anteriormente. Pondo isto, organizações farmacêuticas e investigadores na área da biotecnologia têm trabalhado sinergicamente de modo a proporcionar uma oportunidade única para a comunidade científica e médica aplicarem-no na terapia de células humanas, como a terapia génica, correção de defeitos em genes de células progenitoras, doenças genéticas, tratamento do cancro e mesmo do HIV, HPV e EBV em células infetadas (Charpentier, 2015; Kennedy et al., 2015). No caso do HPV e HBV, a interrupção dos genes virais representa um meio altamente eficaz para a eliminação do vírus das células infetadas. O sistema CRISPR/Cas9 tem demonstrado ser uma técnica bastante eficaz na eliminação dos vírus das células infetadas e na cessação dos tumores (Kennedy et al., 2015; LaFountaine et al., 2015), assim como no tratamento de doenças monogénicas hereditárias (LaFountaine et al., 2015) já estudadas *in vivo* em embriões de primatas não humanos (LaFountaine et al., 2015). Outra vantagem observada em relação às outras técnicas de edição de genoma é o facto de, após induzir uma DSB, o componente proteico Cas9 manter-se inalterado (Gupta & Musunuru, 2014; Sander & Joung, 2014; Vasileva et al., 2015). Esta facilidade de uso da CRISPR/Cas9 demonstrou ser uma vantagem, especialmente na produção de grandes conjuntos de vetores para atingirem locais alvos ou mesmo genomas extensos (Gupta & Musunuru, 2014; Sander & Joung, 2014; Vasileva et al., 2015). Temos também como vantagem o facto deste sistema ser

capaz de modificar múltpiplos locais genómicos simultaneamente, ou seja, de utilizar vários RNAs guia em paralelo dentro da mesma célula (Gupta & Musunuru, 2014; Vasileva et al., 2015). Este facto tornou então mais simples a mutação de múltiplos genes de uma só vez ou a indução e deleções precisas numa região genómica, embora deva de se ter em atenção que o uso em simultâneo de vários pares de ZFN e TALEN pode alcançar os mesmos resultados (Gupta & Musunuru, 2014; Sander & Joung, 2014; Vasileva et al., 2015).

Como todas as técnicas de engenharia genética as desvantagens também são observadas, tais como mutações fora do alvo mesmo nas sequências mais conservadas, sobretudo em células humanas, podendo ser algo preocupante (Gupta & Musunuru, 2014; Mahfouz et al., 2014; Sander & Joung, 2014). Apesar destes contratempos, é possível gerar este sistema CRISPR/Cas9 com alta eficiência e especificidade (Mahfouz et al., 2014). Vários estudos recentes apontam que desajustes na sequência alvo são tolerados pelo sistema. A sua capacidade de os tolerar provem de uma defesa específica das bactérias contra as invasões virais provenientes de DNA transportado pelos plasmídeos, como referido anteriormente (Mahfouz et al., 2014). Recentemente foram feitos dois estudos comparativos sobre a especificidade alvo entre uma proteína Cas9 cataliticamente inativa (Mahfouz et al., 2014; Sander & Joung, 2014) – dCas9- e uma Cas9 nativa, onde se observaram os efeitos fora do alvo de ambas (Mahfouz et al., 2014). A dCas9 apresentou elevadas mutações fora do alvo, principalmente na cromatina, sendo que o PAM proximal desempenhou um papel importante na determinação desta especificidade de ligação. A Cas9 nativa mostrou igualmente algumas mutações fora do alvo, apesar da taxa ser menos elevada. Estes resultados mostraram então a necessidade de se procurar inovações nas variantes das endonucleases das Cas, uma vez que as atividades de clivagem dentro e fora do alvo parecem variar com o tipo de células utilizadas, sendo então bem possível que haja variações entre as espécies eucarióticas usadas (Gupta & Musunuru, 2014; Mahfouz et al., 2014). A Cas utilizada vai depender da espécie eucariótica e do objectivo final, sendo que a normalmente empregue é a Cas9 proveniente de *Streptococcus pyogenes* (Charpentier, 2015; Gupta & Musunuru, 2014; Mahfouz et al., 2014; Sander & Joung, 2014; Vasileva et al., 2015), descoberta em 2013. Apesar de ser a mais comumente utilizada, as diferentes versões da Cas9 possuem necessidades diferentes de PAM, permitindo um direcionamento para locais alvo do genoma aos quais o *S.pyogenes* não consegue alcançar. Para isso, foram desenvolvidas Cas9 obtidas a partir

de *Neisseria meningitidis*, as quais demonstraram ser mais tolerantes a este tipo de variações (Gupta & Musunuru, 2014; Vasileva et al., 2015).

Ainda relativamente às mutações fora do alvo, estas ocorrem normalmente com um tamanho de até cinco a seis pares de bases (LaFountaine et al., 2015), sendo encontradas normalmente quando nos deparamos com sequências PAM alternativas à forma 5'NGG3 a que estamos habituados (Sander & Joung, 2014).

De modo a aumentar a especificidade do sistema em células de mamíferos, houve a necessidade de projetar uma versão mutante da Cas9 que induz apenas corte de cadeia simples no DNA alvo em vez de uma DSB. O uso de um par de nickases de CRISPR/Cas9 com locais de ligação em locais opostos da cadeia de DNA que flanqueiam o local alvo pode produzir o equivalente a uma DSB com saliência em 5', que são posteriormente reparadas por NHEJ ou HDR e pode resultar numa alteração no DNA alvo (Gupta & Musunuru, 2014; Vasileva et al., 2015).

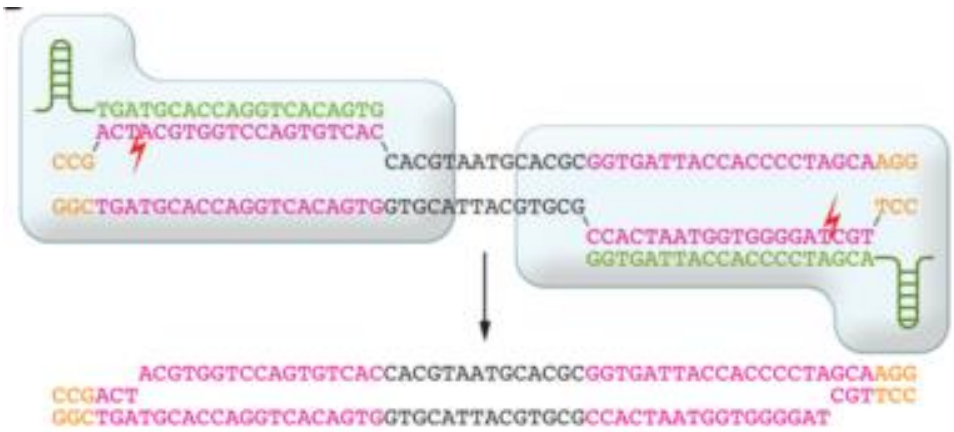

Figura 10: Nickase criada pelo sistema CRISPR/Cas9 de modo a flanquear as sequências de DNA e gerar cortes de cadeia simples que são equivalentes a uma DSB (Gupta & Musunuru, 2014)

Esta alternativa vai reduzir as mutações fora do alvo, uma vez que é pouco provável que uma alteração de uma única cadeia de DNA seja fixada por um mecanismo diferente, como a excisão de bases por via de reparação. Como a probabilidade de duas nickases próximas se ligarem ao resto do genoma é baixa, a taxa de mutações fora do alvo vê-se reduzida (Gupta & Musunuru, 2014).

Outra estratégia para reduzir os efeitos fora do alvo é reduzir o comprimento da porção de RNA guia e respetivo protospacer, o que faz com que seja menos propicio a desarranjos e, assim poder preservar a eficácia no alvo enquanto as mutações fora do alvo são reduzidas. Uma última estratégia bem sucedida é a utilização de um par de proteínas,

cada uma com uma nucleasse Cas9 inativada, ficando incapacitada de clivar o DNA, fundida com um domínio FokI. Cada Cas9 é direcionada para qualquer uma das duas sequências de RNA guia, sendo que o posicionamento da FokI pode formar dímeros e gerar DSBs. Embora de grandes dimensões, estas proteínas de fusão combinam as suas propriedades com as mais desejadas dos sistemas ZFN e TALEN (Gupta & Musunuru, 2014)

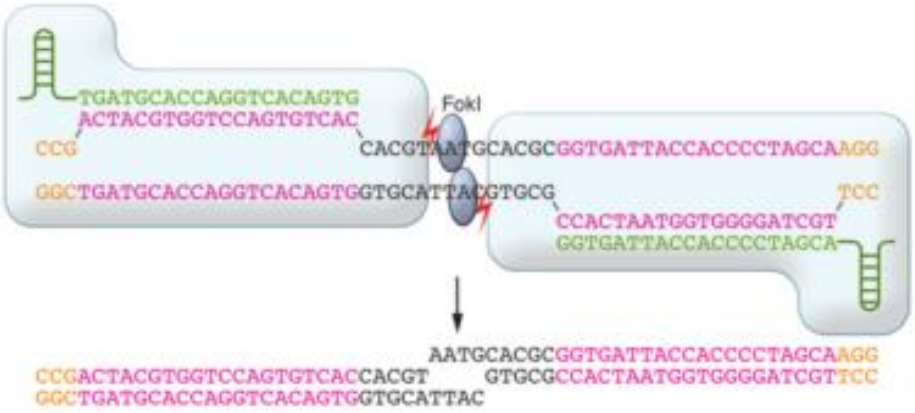

Figura 11: Proteínas de fusão CRISPR/Cas9 e domínios FokI cataliticamente inativos ligam-se a sequências de DNA e posicionam os seus domínios FokI de modo a dimerizar e gerar uma DSB nos locais de ligação (Gupta & Musunuru, 2014)

Relativamente às DSBs, tal como nas técnicas de edição de genoma abordadas anteriormente, também o sistema CRISPR/Cas9 é capaz de as induzir no DNA, neste caso, 3 nucleótidos a montante do PAM. Em contraste com a ZFN e TALEN, que requerem uma recodificação de proteínas utilizando segmentos de DNA grandes, cerca de 500-1500pb, para cada novo local alvo, o sistema CRISPR/Cas9 pode ser facilmente adaptado a qualquer sequência alvo, uma vez que altera 20pb do protospacer do RNA guia (Gupta & Musunuru, 2014; Robert et al., 2015; Vasileva et al., 2015).

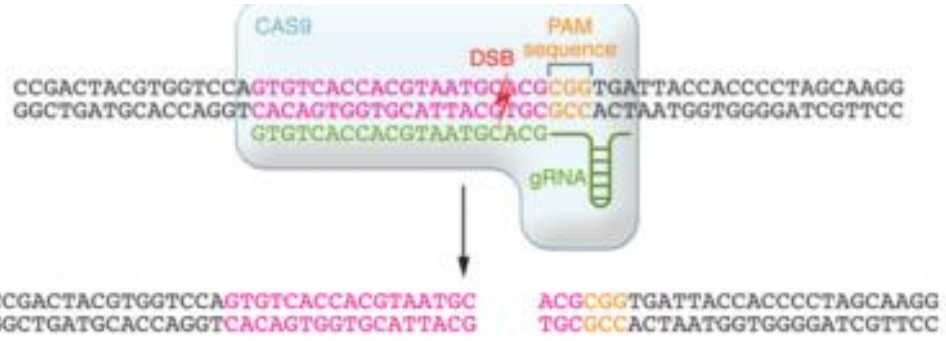

Figura 12: O sistema CRISPR/Cas9 utiliza RNA guia que reconhece e hibrida um protospacer de 20pb no genoma. A DSB ocorre num local 3pb a montante da sequência do PAM (Gupta & Musunuru, 2014)

Estas DSBs podem ser reparadas por NHEJ, dando origem a inserções ou deleções, ou através de HDR. NHEJ é iniciado por uma DSB levando ao recrutamento de um heterodímero e de uma subunidade catalítica de uma proteína quinase dependente de DNA (DNA-PKcs) que se junta a duas extremidades, desencadeando uma série de reações tais como resseção, extensão e ligação (Robert et al., 2015). No que diz respeito a HDR, as DSBs são processadas numa saliência da cadeia simples 3' e um conjunto diferente de proteínas de reparação vão estimular a invasão por um modelo homólogo dador, seguida da reparação desejada (Robert et al., 2015). A identificação das células de interesse editadas por HDR é um processo que envolve uma triagem de clones individuais para poderem ser identificadas posteriormente. Para melhorar a sua eficiência, é necessário reduzir significativamente a carga envolvida na identificação dos clones de interesse a jusante (Robert et al., 2015).

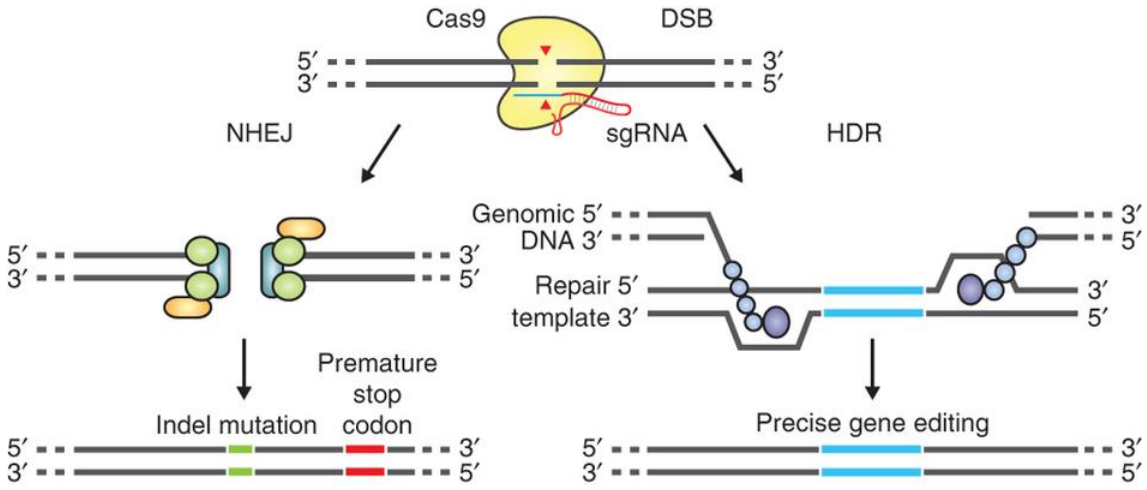

Figura 13: A reparação das DSBs promove a edição dos genes. As DSBs induzidas pelas Cas9 – a amarelo – podem ser reparadas de duas maneiras: via NHEJ ou HDR. NHEJ tem tendência a originar erros, onde as suas extremidades são processadas por máquinas de reparação do DNA endógeno, podendo resultar em mutações aleatórias nos locais de junção. As mutações que ocorrem dentro da região codificadora de um gene podem resultar num quadro de leitura e na criação de um codão prematuro de paragem no gene, resultante na eliminação da sua atividade. Em alternativa podemos utilizar o modelo de reparação HDR, onde não existe o risco de mutações e pode ser alcançada uma edição precisa do gene (Ran et al., 2013)

Geralmente, a via NHEJ representa um meio de reparação mais frequente quando se utiliza uma engenharia mediada por Cas9, mesmo estando na presença de um modelo dador especifico para HDR (Robert et al., 2015). Este facto é apoiado também por o método de NHEJ ocorrer ao longo de todo o ciclo celular, enquanto que HDR é uma

reparação que necessita de ser mediada pela fase G2 (Robert et al., 2015). O método de reparação NHEJ apresenta níveis mais elevados de linhas celulares com deleções ou inserções do que o método de HDR, sugerindo então uma competição entre estes dois processos (Robert et al., 2015).

Como referido anteriormente, este sistema de edição de genoma é uma ferramenta poderosa para a regulação de genes alvo endógenos. Apesar de todas as suas aplicações, os investigadores viram-se na necessidade de aumentar o seu alcance e oferecer um maior controlo do espaço-tempo da expressão do gene e assim criar uma ferramenta optogénica baseada nesta proteína, permitindo a ativação dos mesmos genes endógenos através da estimulação por uma luz (Nihongaki et al., 2015). Esta técnica pode também ser alcançada através do sistema LITE abordado no capítulo das TALENs, apesar de que a substituição das TALE pelas proteínas Cas9 não proporciona a mesma taxa de ativação conseguida pelo método baseado nas dCas9 óticas (Nihongaki et al., 2015). Este facto pode ser explicado por a dCas9 possuir uma estrutura maior e mais complexa que as TALEN, podendo assim enfraquecer mais facilmente a localização e induzir a dimerização das duas proteínas de fusão. Sendo assim, foi desenvolvido um sistema de ativação baseado neste gene dCas9 ótico, que pode permitir a ativação de genes alvo de células de mamíferos de modo rápido e reversível, assim como uma engenharia genómica e aplicações biotecnológicas (Nihongaki et al., 2015).

O desenvolvimento deste sistema de fotoativação do genoma baseado na dCas9 e uma cryptochrome 2 sensível à luz – CRY2 – e a sua ligação a CIB1 de *Arabidopsis thaliana* veio oferecer um modo simples e versátil de ativar múltiplos genes alvo (Nihongaki et al., 2015).

O sistema fotoativável a partir do sistema CRISPR/Cas9 consiste em duas proteínas de fusão que se ligam ao sgRNA (Nihongaki et al., 2015). A primeira proteína de fusão é uma sonda genómica que contêm a dCas9 e a CIB1, funcionando como uma âncora que se liga à sequência alvo do genoma guiada pelo sgRNA. A segunda proteína de fusão é o ativador da sonda, que inclui uma região de fotólise homóloga de CRY2 - CRY2PHR – e o domínio ativador da transcrição.

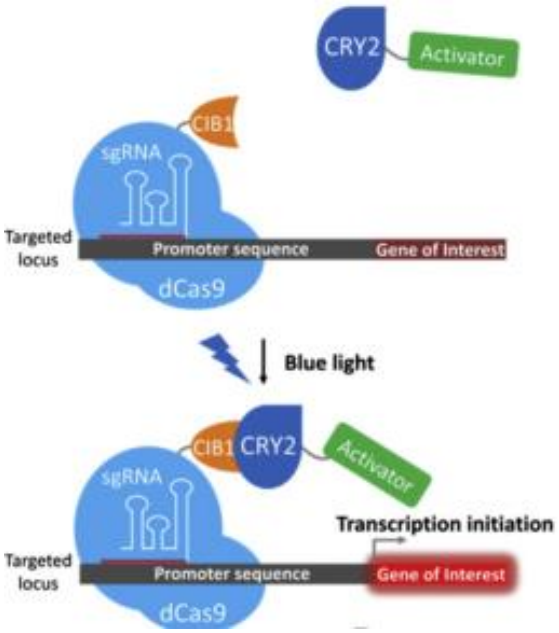

Figura 14:Esquema da transcrição baseada no sistema fotoativável CRISPR/Cas9 (Nihongaki et al., 2015)

Na ausência de luz azul, a âncora liga-se à região promotora do genoma alvo por meio do sgRNA, enquanto que a sonda ativadora é difundida livremente no interior do núcleo. Na presença de luz azul, CRY2PHR e CIB1 são heterodimerizados e, consequentemente, o domínio ativador da transcrição é recrutado para o local de destino para ativar a expressão do gene alvo (Nihongaki et al., 2015). Foram desenvolvidos vários estudos baseados neste sistema, concluindo que todas as amostras expostas a esta luz apresentaram expressão do gene e que quanto maior o tempo de exposição, maior a expressão do gene (Nihongaki et al., 2015).

O tamanho da proteína Cas9 é outro problema com o qual os investigadores se têm vindo a debater. A região codificante do cDNA do *S.pyogenes* é de aproximadamente 4,2 kb, tornando-o um pouco maior do que um monómero de TALEN e muito maior que o de ZFN, apesar de estes dois serem alvo de dimerização, o que os torna maiores, mas mesmo assim nunca tão grandes como CRISPR/Cas9 (Gupta & Musunuru, 2014; Vasileva et al., 2015). Este tamanho torna difícil a sua entrega através de vetores virais, que requerem um promotor e uma sequência adicionais. O RNA guia tem cerca de 100pb e o promotor, normalmente o U6, 500pb (Gupta & Musunuru, 2014; Vasileva et al., 2015).

Apesar dos enormes avanços nesta técnica, a entrega eficiente deste sistema continua a ser um enorme desafio (Kennedy et al., 2015). O vetor viral AAV, usado normalmente para a transdução de tecidos de tumores de fígado, neurónios, músculos e tecidos de tumor in vivo (Kennedy et al., 2015), não é o mais indicado para esta via devido ao seu tamanho, 4,5kb, limitando assim a vectorização do gene spCas9 com carga aproximada de 4,2 kb, demasiado grande para que este vetor seja expressado e que consiga expressar o gene e o sgRNA correspondente (Kennedy et al., 2015; LaFountaine et al., 2015; Sander & Joung, 2014; Vasileva et al., 2015). Apesar destas aparentes limitações, os investigadores têm-se esforçado por construir estes vetores com citomegalovírus, de modo a incorporar eficazmente a proteína Cas9 e o sgRNA (LaFountaine et al., 2015). Para reduzir o seu tamanho global, têm sido desenvolvidos sinais de terminação mínimos e sequências promotoras que não afetam as frequências de edição. Além disso, demonstrou-se que diferentes promotores podem ser introduzidos para atingir linhas celulares específicas *in vivo* e *in vitro* (LaFountaine et al., 2015).

Alternativamente, outros investigadores desenvolveram vetores AAV separados, um para Cas9 e outro para sgRNA, com promotores específicos, de modo a aliviar as restrições de tamanho da sua cápsula (LaFountaine et al., 2015). Foi também descoberto que uma proteína Cas9 de menor tamanho, cerca de menos 1kb que a normalmente usada de *S. pyogenes,* proveniente do *Staphylococcus aureus* – SaCas9 – é mais fácil de agregar em vetores AAV (LaFountaine et al., 2015). Este novo desenvolvimento foi testado em ratinhos, onde se observou uma diminuição do seu colesterol total. Esta eficiência foi observada quando se utilizou a enzima Cas9 mais pequena para integrar nos vetores AAV (LaFountaine et al., 2015).

De modo a entregar sistemas CRISPR/Cas9 maiores, os lentivírus parecem ser os mais indicados para inserir o sgRNA, melhorando bastante a edição múltipla de genomas, onde são necessários vários gRNA diferentes em simultâneo (Kennedy et al., 2015; LaFountaine et al., 2015; Sander & Joung, 2014; Vasileva et al., 2015).

A indecisão entre qual dos vetores utilizar é significativa, uma vez que os lentivírus apenas são capazes de acomodar o sistema CRISPR/Cas9 proveniente de *S.pyogenes,* enquanto que os AAV não os conseguem acomodar devido ao seu tamanho limitado de cerca de 5kb (Gupta & Musunuru, 2014). Devido a este facto foi novamente necessário refletir sobre a disponibilidade de obter sistemas CRISPR/Cas9 a partir de outras espécies (Gupta & Musunuru, 2014). Como exemplo temos a *Neisseria*

meningitidis, com um tamanho mais reduzido, de aproximadamente 3,2 kb, permitindo assim a entrega através de AAV (Gupta & Musunuru, 2014).

A entrega não viral dos plasmídeos contendo Cas9 tem sido estudada, relatando-se o uso de lípidos catiónicos ou de nanopartículas poliméricas (LaFountaine et al., 2015). Péptidos com capacidade de penetrar em células têm sido igualmente estudados com a finalidade de promoverem a absorção celular do sistema CRISPR/Cas9, embora seja complicado devido à necessidade de introduzir na célula uma proteína e um RNA (LaFountaine et al., 2015).

Temos igualmente de ter em conta que a expressão dupla do sgRNA tem aumentado significativamente o potencial destes sistemas de edição (Kennedy et al., 2015). A clivagem de múltiplos locais é igualmente crucial para eliminar mutações individuais do sgRNA *in vivo*. Além disso, um vetor que transporta um sgRNA pode facilmente ser adaptado para nickases por adição de uma nickase Cas9, podendo aumentar a especificidade e assim clivar o DNA alvo (Kennedy et al., 2015). A entrega de um sgRNA multiplex Cas9 e o uso de vetores AAV pode então ter o potencial para prolongar consideravelmente a utilização do sistema CRISPR/Cas9 *in vivo* (Kennedy et al., 2015).

Este sistema de reparação tem demonstrado ser eficaz para a reparação e edição de genes e assim como uma grande ajuda no combate a certas doenças.

Anteriormente foi abordado o facto desta técnica estar a ser estudada de modo a alcançar o tratamento de doenças monogénicas hereditárias, tendo como exemplo a anemia falciforme, DMD e A1ATD (LaFountaine et al., 2015), já testadas também para as outras duas técnicas de edição de genoma. Os resultados obtidos para as suas eficiências foram diferentes, sendo que o sistema CRISPR/Cas9 demonstrou cerca de mais de 100 inserções e/ou deleções, comparativamente com a técnica TALEN (LaFountaine et al., 2015) em células hiPSCs. Esta comparação foi realizada através da utilização de NHEJ e HDR em ambos os casos (LaFountaine et al., 2015).

A edição de genes através deste sistema tem sido igualmente usado para corrigir desordens genéticas em ratinhos que sofrem de cataratas (LaFountaine et al., 2015). Uma mutação dominante no gene *Crygc* foi corrigida através de uma co-injeção de mRNA de Cas9, um gRNA e um oligonucleótido dador de zigotos ou células estaminais espermagotoniais (SCC) (LaFountaine et al., 2015). No primeiro caso, alguns embriões fertilizados desenvolveram fenótipos normais e a mudança foi igual para toda a linha germinativa. No caso das SCC, foram produzidos descendentes saudáveis com uma eficiência de 100% por meio da geração de espermatídeos corrigidos usados na

fertilização (LaFountaine et al., 2015). Um procedimento semelhante é previsto para as SSCs em humanos para tratar doenças dominantes em alelos paternos e doenças genéticas ligadas ao cromossoma X (LaFountaine et al., 2015). A comprovação desta hipótese foi também observada para terapias antivirais mediadas tanto por sistemas CRISPR/Cas9 como para ZFN e TALEN (LaFountaine et al., 2015).

Semelhantemente ao observado para ZFN e TALEN, o sistema CRISPR/Cas9 pode igualmente, de modo eficiente, induzir a inativação do gene CCR5, de modo a impedir a integração do vírus HIV1 nas células (LaFountaine et al., 2015). Foi utilizado com a finalidade de integrar de um modo direcionado o genoma de HIV1 diretamente nas células humanas, erradicando assim o vírus nas células testadas. Este direcionamento direto de sequências do genoma viral sem homologia relacionada com a do genoma hospedeiro, é uma estratégia que também tem sido introduzida para HPV e EBV, de modo a tratar igualmente estas infeções virias e reduzir possíveis mutações no genoma do DNA humano (LaFountaine et al., 2015).

Observou-se que este sistema foi igualmente eficaz na correção da mutação do gene CFTR, responsável pela fibrose cística, em células estaminais intestinais de doentes. Organóides doentes foram transfetados com sgRNAs que tiveram como alvo o intrão 11 e o exão 11 do gene CFTR, juntamente com um plasmídeo que codifica a sequência dadora de tipo selvagem de CFTR. Vários organoides que continham o tipo selvagem deste gene foram isolados.(LaFountaine et al., 2015). Apesar de estes resultados serem promissores, múltiplos órgãos estão afetados pela fibrose cística, não o tornando um candidato atraente para terapias baseadas em células estaminais induzidas por este sistema CRISPR/Cas9 (LaFountaine et al., 2015).

A edição de genoma a partir deste sistema tem sido avaliada como uma potencial alternativa à terapia com estatinas para prevenir a doença cardíaca coronária. A proteína PCSK9 expressa nas células hepáticas, atua como um antagonista de lipoproteínas de baixa densidade (LDL), resultando em níveis mais elevados de colesterol LDL. Indivíduos com uma mutação nesta proteína apresentam cerca de 30 a 80% de redução nas concentrações de LDL, desenvolvendo efeitos adversos. Assim, a inativação do gene que codifica para esta proteína foi realizada em ratinhos através da entrega de CRISPSR/Cas9 por adenovírus (LaFountaine et al., 2015; Mahfouz et al., 2014). Quatro dias após a inativação da proteína, as mutações foram tão elevadas que chegaram aos 50% nas células do fígado, reduzindo em cerca de 40% os níveis de colesterol no plasma (LaFountaine et al., 2015). Isto sugere então que o sistema CRISPR/Cas9 pode se usado

como um tratamento único, podendo conduzir a um efeito terapêutico que oferece uma vantagem substancial sobre as terapias baseadas em estatinas que devem de ser tomadas cronicamente, resultando por vezes em baixa aderência por parte do doente (LaFountaine et al., 2015).

Além de permitir uma edição de genoma de forma fácil e eficiente, o sistema CRISPR/Cas9 tem a vantagem de ser utilizado para regular a expressão de genes endógenos e marcar locus específicos do cromossoma em células e organismos vivos (Sander & Joung, 2014).

Células estaminais e técnicas de edição de genoma

Como referido anteriormente, o aparecimento das técnicas de edição de genoma tem vindo a desenvolver-se ao longo dos últimos anos, sendo altamente benéficas para o estudo de modelos de doenças humanas (Vasileva et al., 2015).

Resultados recentes na produção de células estaminais de pacientes e a sua diferenciação específica em vários tipos de células somáticas facilitou grandemente estes desenvolvimentos acerca da fisiopatologia de doenças socialmente importantes (Vasileva et al., 2015).

As células estaminais incluem as células estaminais embrionárias (ESC) e as células estaminais pluripotentes induzidas (iPSCs). As células estaminais são capazes de se proliferar indefinidamente, auto renovar-se e de se transformar em linhas de células diferenciadas que oferecem uma oportunidade de produção de modelos de doenças humanas (Vasileva et al., 2015).

As células estaminais pluripotentes humanas fornecem uma plataforma versátil de estudos regenerativos, incluindo testes de drogas e produção de modelos de doenças. A expressão de apenas 4 fatores de transcrição – Oct4 (é o mais importante de entre todos), Klf4, Sox2 e Myc (OSKM)– é suficiente para a geração de iPSCs a partir de células somáticas diferenciadas. Estas células revolucionaram o campo da genética e também puseram em evidência a importância da reprogramação por OSKM como meio de edição de genoma, assim como a sua capacidade de superar os traços epigenéticos de células diferenciadas e reverte-los para o estado de pluripotência inicial (Vasileva et al., 2015). A pluripotência destas células é caracterizada por uma configuração específica da cromatina e consequentes modificações epigenéticas (LaFountaine et al., 2015)

A atividade e expressão dos quatro fatores de transcrição, denominados OKSM, é reprimida em células somáticas normais e, portanto, a sua reativação é fundamental para a reprogramação de células somáticas em células estaminais. A sua transferência direta através de cDNAs em células somáticas não pode ser utilizada para efeitos de terapia genética devido à probabilidade de recombinação indesejada com o DNA (Vasileva et al., 2015).

Uma alternativa ao uso dos OSKM é a reativação das células diferenciadas, podendo esta ser alcançada através de um direcionamento específico de domínios de transcrição por meio de técnicas de edição de genoma, como ZFNs, TALENs e CRISPR/Cas9 (Vasileva et al., 2015), permitindo o sucesso de geração de células estaminais pluripotentes induzidas – iPSCs- (LaFountaine et al., 2015).

A fusão da CRISPR/Cas9 com o domínio da histona da acetiltransferase de p300 pode reativar, por si própria, o locus silenciado epigeneticamente do Oct4, o que torna este sistema uma ferramenta bastante atrativa para a produção de iPSCs (LaFountaine et al., 2015).

A produção de células iPSCs requer a inativação da proteína antitumoral p53, que, por sua vez, provoca a tumorogénese, sendo interessante ver o modo como a Cas9 se comporta temporariamente perante uma inibição mediada por 3pb alvo a jusante, sendo que a proteína p53 não é suficiente para desencadear uma desdiferenciação sem afetar o controlo de qualidade da mesma (LaFountaine et al., 2015; Vasileva et al., 2015)

O sistema CRISPR/Cas tem sido aplicado com sucesso para a edição de genoma de plantas, insetos, peixes, ratos, ratinhos, linhas de células humanas CES e iPSCs. Mais recentemente, a partir de uma nova plataforma – iCRISPR- baseada nos sistemas CRISPR/Cas e TALENs, foi projetado para a rápida e eficiente produção de bi-alelos knockout em linhas de células hPSCs. Em primeiro lugar, a geração destas linhas de células hPSCs que expressam Cas9, o componente invariável do sistema CRISPR/Cas, foi criada para a realização de tal plataforma (Vasileva et al., 2015).

Para o próximo passo, a transfeção de pequenos RNAs mediada por lípidos, foi determinada como eficiente para a co-transfeção de vários gRNAs para a edição simultânea de múltiplos alvos de genoma durante a fase desejada de hPSC. Para tornar a plataforma iCRISPR mais flexível e ajudar na expressão, foram produzidas formas do gene Cas9 indutíveis por doxiciclina através de TALENs. Esta plataforma permite a geração sucessiva de duplos e triplos knockouts de linhas de hPSCs, bem como uma indução especifica do estadio em que o gene irá sofrer knockout durante a diferenciação das hPSCs. Além disso, o sistema CRISPR com nucleases Cas9 mortas (dCas9) funde-se com um domínio de ativação de transcrição. A ativação e a repressão de genes específicos em hPSCs pode afetar o curso de diferenciação que tem sido alcançado através da utilização deste sistema (Vasileva et al., 2015).

HIV e técnicas de edição de genoma

A SIDA continua a ser um problema incurável de saúde pública, devido à integração do vírus HIV-1 no genoma humano, concedendo um risco de reativação viral, mesmo depois da administração da terapia adequada (Hu et al., 2014). A terapia passa por uma administração oral de uma mistura de compostos anti-retrovirais que inibem as várias etapas do ciclo de vida viral, reduzindo profundamente a sua replicação nas células do hospedeiro e reduzindo a infeção plasmática a nível mínimo, no entanto não são capazes de suprimir totalmente a expressão do genoma viral e a sua consequente replicação nos tecidos e células alvo que servem de reservatório para este vírus (Hu et al., 2014)

Ao longo de vários anos foram desenvolvidas novas estratégias com a esperança de erradicar este vírus totalmente, de modo a conseguir retirar o genoma viral a partir de células ocultas infetadas, pois os métodos atuais têm-se mostrado ineficientes e propensos a vários efeitos fora do alvo (Hu et al., 2014).

Como foi referido ao longo deste trabalho, inicialmente utilizaram-se as técnicas de ZFN e TALEN com a esperança de perturbar estes genes e erradicar seletivamente o DNA viral do HIV-1 do genoma do hospedeiro (Hu et al., 2014; Khalili et al., 2015). Durante alguns anos foram então utilizadas como alvos terapêuticos os seus co-recetores, tais como o CCR5 e o CXCR4, de modo a interromper a entrada do vírus na célula (Hu et al., 2014; Khalili et al., 2015).

O sistema CRISPR/Cas9 é a técnica de edição de genoma mais recentemente utilizada, na esperança de alcançar a eliminação do vírus HIV-1, através de configurações individuais ou multiplex. A sua alta eficiência de alcance para o alvo, rapidez de produção e boa transmissão da linha germinativa tornaram este sistema tão interessante para ser aplicação na procura da cura para a SIDA (Khalili et al., 2015). Outro aspeto a ter em conta é o facto de o sistema CRISPR/Cas9 induzir uma melhor eficiência de clivagem – 33%- em relação às TALEN (14%) (Khalili et al., 2015). Como referido anteriormente, foram identificados alvos altamente específicos dentro da região LTR U3, tendo estes sido eficientemente editados por esta técnica de edição de genoma (Hu et al., 2014). Observou-se uma inativação da expressão do gene viral e a sua consequente não replicação em microglia , promielóctios, astócitos, células linfóides e células T CD4+ (Hu et al., 2014; Khalili et al., 2015).

O uso dos gRNAs não causa toxicidade nem efeitos fora do alvo consideráveis nas células hospedeiras, erradicando completamente um fragmento de 9709 pb de DNA viral integrado que se estende a partir da extremidade 5' até 3' de terminações longas repetitivas – LTR (Hu et al., 2014; Khalili et al., 2015). As gRNA/Cas9 dirigidas aos LTR vão erradicar o genoma do HIV-1 e imunizar eficazmente as células contra este vírus e as suas possíveis reativações (Hu et al., 2014; Khalili et al., 2015). Além disso, a presença dos gRNAs multiplex no interior das células que expressam Cas9, impediu a infeção por HIV-1 (Khalili et al., 2015). Os potenciais efeitos fora do alvo causados por esta técnica não são uma grande preocupação devido à baixa homologia observada entre o genoma viral e o genoma endógeno do hospedeiro (Khalili et al., 2015)

O mais interessante no uso desta técnica é a sua capacidade de eliminar mais do que uma cópia do genoma viral integrado na célula infetada, mesmo quando o vírus está em simultaneamente em diferentes regiões do genoma hospedeiro (Khalili et al., 2015).

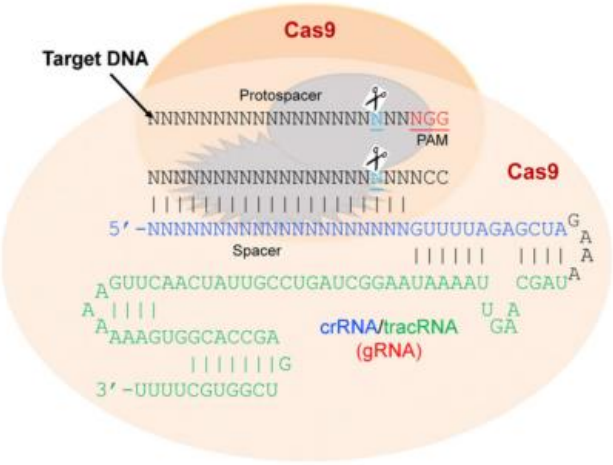

Figura 15: Representação da clivagem precisa do terceiro nucleótido do PAM pelo complexo RNA-Cas9. Este induz DSBs no DNA alvo. Podemos ver representado o gRNA que é uma agregação do crRNA e do tracRNA(Hu et al., 2014)

Na imagem acima apresentada podemos observar que o sistema é capaz de alterar as sequências de DNA em ambas as cópias de DNA viral, podendo então ser capaz de alterar sequências de DNA viral em células infetadas ocultas que contêm cópias deste mesmo DNA (Khalili et al., 2015).

Estudos desenvolvidos mostraram então que os gRNAs usados na técnica de CRISPR/Cas9 podem constituir uma abordagem terapêutica alternativa mais indicada

para a profilaxia da SIDA (Hu et al., 2014), uma vez que as três técnicas foram testadas quanto à sua eficiência na interrupção do gene CCR5 e CXCR4 (Khalili et al., 2015).

Conclusão e perspetivas futuras

Como abordado ao longo de todo o trabalho, as técnicas de edição de genoma estão focadas em tratamentos revolucionários e de sucesso para o cancro, doenças monogénicas, doenças infeciosas (HIV, HBV, HPV) e doenças cardiovasculares (LaFountaine et al., 2015). Foi demonstrado através de vários estudos que o sistema imunitário dos doentes pode ser treinado para alcançar e destruir as células tumorais através das técnicas abordadas ao longo do trabalho (LaFountaine et al., 2015).

Apesar dos seus enormes desenvolvimentos nos últimos anos, muitas são as perguntas que surgem frequentemente. Há métodos já experimentados mas que necessitam de ser continuamente testados e melhorados de modo a obter uma técnica o mais eficaz possível.

Os estudos iniciais basearam-se nestas técnicas de modo a solucionar questões tais como o tratamento de doenças monogénicas hereditárias e infeções virais *in vivo* e *ex vivo*, dependendo da perturbação ou do tecido afetado (LaFountaine et al., 2015; Sander & Joung, 2014).

A administração *ex vivo* consiste no isolamento e cultura de linhas celulares a partir de um indivíduo inicialmente infetado, seguido de uma correção genética *in vitro*. As células são corrigidas geneticamente e administradas *in vivo* de volta no doente. Estas células podem-se multiplicar ainda mais *in vivo*, e assim restaurar o fenótipo do doente, parcial ou totalmente (LaFountaine et al., 2015). Esta abordam permite um pré rastreamento das células para possíveis mutações fora do alvo que podem conduzir a efeitos adversos e um nível muito mais elevado de transfeção e correção genetética (LaFountaine et al., 2015). Esta abordagem só é viável para tecidos e células que são de fácil acesso e proliferativas, tais como as do sangue, pele e músculos (LaFountaine et al., 2015).

Recentes avanços visam o uso de hiPSCs, tendo aberto ainda mais portas para o contínuo uso das terapias genéticas *ex vivo* (LaFountaine et al., 2015). A capacidade de tomar uma amostra de tecido humano e reprogramá-lo em células estaminais que depois podem ser diferenciadas em qualquer outro tipo de célula humana, influenciou a sua quase obrigatoriedade em terapia genética *ex vivo* (LaFountaine et al., 2015).

Recentemente, o sistema CRISPR/Cas9 foi usado para modificar uma linha de células estaminais e hematopoiéticas. Isto ilustra o grande potencial do sistema para a medicina personalizada, uma vez que é capaz de modificar linhas de células estaminais (LaFountaine et al., 2015).

Outro grande avanço deste sistema foi a sua capacidade de criar órgãos *in vitro*, corrigidos para qualquer erro genético que o indivíduo possa ter, abriu infinitas possibilidades de tratamento e criou uma enorme esperança a nível médico (LaFountaine et al., 2015). A redução da necessidade de recorrer a ensaios animais e humanos vê-se igualmente diminuída com este desenvolvimento. Apesar de ser bastante dispendiosa, as pesquisas desenvolvidas com base nesta abordagem demonstraram que estão abertas as portas para uma medicina personalizada (LaFountaine et al., 2015).

Como referido anteriormente, a terapia *ex vivo* está mais direcionada para o tratamento de células como as do sangue, pele e músculo, enquanto que a *in vivo* permite uma terapia mais direcionada para células do fígado e cérebro, por exemplo, através da entrega mediada por vetores virais ou não virais, dependendo do pretendido (LaFountaine et al., 2015).

Os métodos de ampliação da gama de clivagem de Cas9 através de gRNA são importantes para induzir HDR precisas e NHEJ, assim como para a implementação de estratégias multiplex, incluindo nickases emparelhadas (Sander & Joung, 2014).

Relativamente à NHEJ e HDR, é necessário resolver o problema relativo ao desequilíbrio existente no que diz respeito ao tamanho das mutações induzidas por NHEJ e as alterações orientadas por HDR, de modo a garantir o sucesso prentendido (Sander & Joung, 2014).

Embora possam ser alcançadas altas taxas de HDR através do sistema CRISPR/Cas9 e uma cadeia simples de oligonucleótidos de DNA, é sempre possível a ocorrência competitiva de uma mutação induzida por NHEJ. Esta competição é vista por muitos investigadores como um problema, uma vez que quando é utilizado o sistema HDR para induzir mutações pontuais no protospacer alvo, muitos alelos que tenham sido alterados podem sofrer novas alterações pelo sistema NHEJ, reduzindo deste modo o rendimento dos alelos corretamente editados (Sander & Joung, 2014). Apesar dos esforços para inibir NHEJ em aplicações terapêuticas em que HDR é o sistema principal, reduzindo ou eliminando a NHEJ competitiva, podem surgir problemas, uma vez que a

sua inibição pode ser mal tolerada pelas células, dado o seu papel central na reparação do ADN (Sander & Joung, 2014).

Como foi observado ao longo do trabalho, o direcionamento das Cas9, nickases emparelhadas e fusões de Cas9 apresentam algumas restrições, principalmente pela necessidade de uma sequência de PAM correspondente a NGG (Sander & Joung, 2014). É necessário então continuar a explorar alternativas, tais como PAMs direcionados para NAG ou NNGG e o modo como estas sequências são reconhecidas e cortadas (Sander & Joung, 2014). Outras plataformas gRNA:Cas9 com sequências PAM isoladas de espécies diferentes das que estamos habituados a trabalhar, como de *Streptococcus thermophilus*, *Neisseria meningitidis* e *Treponema denticola* têm começado a ser exploradas de modo a aumentar ainda mais o direcionamento destas plataformas (Sander & Joung, 2014).

Outra preocupação que necessita de ser rapidamente ultrapassada é a questão das mutações fora do alvo causadas pelas nucleases Cas9 e as nickases emparelhadas que ocorrem nos genomas alvo (Sander & Joung, 2014). Estão a ser desenvolvidas estratégias cruciais capazes de avaliar a eficácia e especificidade destas plataformas, sendo que o ideal era uma combinação com outras abordagens já existentes (Sander & Joung, 2014).

Exemplos de tais melhorias envolvem a utilização de proteínas de engenharia capazes de modificar a proteína Cas9 e os nucleótidos utilizados pelo gRNA que medeiam o reconhecimento dos locais alvo (Sander & Joung, 2014).

Um desafio relacionado com os métodos de entrega eficiente deste sistema é a otimização para cada tipo de célula particular ou organismo a ser modificado, assim como o desenvolvimento de métodos que permitam a expressão quer de gRNAs, quer de Cas9 específicas para um dado tecido ou tipo de célula (Sander & Joung, 2014). gRNAs devem ser expressos em simultâneo a partir de um mesmo vetor, permitindo assim a uma maior capacidade multiplex destes sistemas (Sander & Joung, 2014).

Tal como relatado anteriormente, o sistema CRISPR/Cas9 representa um método ainda em ascensão, capaz de gerar modelos de ratinhos e neles induzir alterações de alelos, knokouts, entre outras estratégias de melhoramento de perturbações genéticas (LaFountaine et al., 2015).

Pensa-se que a facilidade com que são criados modelos de mamíferos através desta tecnologia vai permitir o crescimento da capacidade de investigar quais as modificações genéticas envolvidas nas doenças humanas (LaFountaine et al., 2015). Isto é apoiado pela já criação de primatas com mutações genéticas específicas, aumentando assim todas as esperanças para o seu uso em humanos (LaFountaine et al., 2015).

No futuro, o uso de métodos de alto rendimento que permitem um perfil global de mutações fora do alvo, deverá proporcionar informações sobre a necessidade do reconhecimento do alvo pertencente a cada sistema (Gaj et al., 2013).

Um dos principais obstáculos que tem de ser rapidamente ultrapassado é a questão da experimentação humana destes sistemas (Mahfouz et al., 2014). A luta por este direito por parte dos investigadores ainda está longe de ser vencida (Mahfouz et al., 2014).

Um dos passos mais importantes a ser dado no futuro é a imunização das células contra o vírus HIV-1 através da produção de vacinas preventivas através da técnica CRISPR/Cas9 (Hu et al., 2014). Estudos comprovaram que a técnica CRISPR/Cas9 leva à eliminação rápida do vírus HIV-1 mesmo antes de ele integrar o genoma do hospedeiro, sendo então uma alternativa altamente viável a ser aplicada no campo da vacinação em várias células latentes e modelos animais *in vivo* (Hu et al., 2014).

Muitos são os esforços para que no futuro a precisão e eficiência do sistema CRISPR/Cas9 continue a aumentar e a melhorar. (Lee, Grav, Lewis, & Kildegaard, 2015). Como referido várias vezes ao longo deste trabalho, as mutações fora do alvo são a principal preocupação dos investigadores e uma das principais correções a alcançar(Lee et al., 2015).

Bibliografia

Anderson, W. F. (1990). Applications o f the P o l y m e r a s e C h a i n Reaction in Retroviral-Mediated G e n e Transfer a n d the Analysis o f G e n e - M a r k e d H u m a n T I L Cells, *149*, 135–149.

Carey, B. W., Markoulaki, S., Hanna, J. H., Faddah, D. a, Buganim, Y., Kim, J., … Jaenisch, R. (2011). Reprogramming factor stoichiometry influences the epigenetic state and biological properties of induced pluripotent stem cells. *Cell Stem Cell*, *9*(6), 588–98. http://doi.org/10.1016/j.stem.2011.11.003

Charpentier, E. (2015). bacterial RNA-guided mechanism opened and biomedicine. *EMBO Molecular Medicine*, *7*(4), 363–365. http://doi.org/DOI 10.15252/emmm.201504847

Chen, K., & Gao, C. (2013). TALENs: Customizable Molecular DNA Scissors for Genome Engineering of Plants. *Journal of Genetics and Genomics*, *40*(6), 271–279. http://doi.org/10.1016/j.jgg.2013.03.009

Cho, S. W., Kim, S., Kim, Y., Kweon, J., Kim, H. S., Bae, S., & Kim, J. (2014). Analysis of off-target effects of CRISPR Cas-derived RNA-guided endonucleases and nickases sup2, 132–141. http://doi.org/10.1101/gr.162339.113.Freely

Cox, D. B. T., Platt, R. J., & Zhang, F. (2015). Therapeutic genome editing: prospects and challenges. *Nature Medicine*, *21*(2), 121–131. http://doi.org/10.1038/nm.3793

De Pater, S., Neuteboom, L. W., Pinas, J. E., Hooykaas, P. J. J., & Van Der Zaal, B. J. (2009). ZFN-induced mutagenesis and gene-targeting in Arabidopsis through Agrobacterium-mediated floral dip transformation. *Plant Biotechnology Journal*, *7*(8), 821–835. http://doi.org/10.1111/j.1467-7652.2009.00446.x

Dougherty, W. G., & Parks, T. D. (1995). Transgenes and gene suppression: Telling us something new? *Current Opinion in Cell Biology*, *7*, 399–405. http://doi.org/10.1016/0955-0674(95)80096-4

Gaj, T., Gersbach, C. a., & Barbas, C. F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in Biotechnology*, *31*(7), 397–405. http://doi.org/10.1016/j.tibtech.2013.04.004

Gupta, R. M., & Musunuru, K. (2014). Expanding the genetic editing tool kit : ZFNs , TALENs , and CRISPR-Cas9. *The Journal of Clinical Investigation*, *124*(10), 4154–4161. http://doi.org/10.1172/JCI72992.transcription

Hu, W., Kaminski, R., Yang, F., Zhang, Y., Cosentino, L., Li, F., … Khalili, K. (2014). RNA-directed gene editing specifically eradicates latent and prevents new HIV-1 infection. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(31), 11461–11466. http://doi.org/10.1073/pnas.1405186111

Joung, J. K., & Sander, J. D. (2013). TALENs: a widely applicable technology for targeted genome editing. *Nature Reviews. Molecular Cell Biology*, *14*(1), 49–55. http://doi.org/10.1038/nrm3486

Kennedy, E. M., Kornepati, A. V. R., Mefferd, A. L., Marshall, J. B., Tsai, K., Bogerd, H. P., & Cullen, B. R. (2015). Optimization of a multiplex CRISPR/Cas system for use as an antiviral therapeutic. *Methods*, 8–12. http://doi.org/10.1016/j.ymeth.2015.08.012

Khalili, K., Gordon, J., Kaminski, R., Cosentino, L., & Hu, W. (2015). Genome editing strategies: potential tools for eradicating HIV-1/AIDS. *Journal of NeuroVirology*. http://doi.org/10.1007/s13365-014-0308-9

Kieffer, B. L. (1999). Opioids: First lessons from knockout mice. *Trends in Pharmacological Sciences*, *20*(January), 19–26. http://doi.org/10.1016/S0165-6147(98)01279-6

LaFountaine, J. S., Fathe, K., & Smyth, H. D. C. (2015). Delivery and Therapeutic Applications of Gene Editing Technologies ZFNs, TALENs, and CRISPR/Cas9. *International Journal of Pharmaceutics*. http://doi.org/10.1016/j.ijpharm.2015.08.029

Lee, J. S., Grav, L. M., Lewis, N. E., & Kildegaard, H. F. (2015). CRISPR/Cas9-mediated genome engineering of CHO cell factories: Application and perspectives. *Biotechnology Journal*, 1–16. http://doi.org/10.1002/biot.201500082

Mahfouz, M. M., Piatek, A., & Stewart, C. N. (2014). Genome engineering via TALENs and CRISPR/Cas9 systems: challenges and perspectives. *Plant Biotechnology Journal*, *12*(8), 1006–1014. http://doi.org/10.1111/pbi.12256

Mandal, A. (n.d.-a). Gene Therapy Issues, 1–2.

Mandal, A. (n.d.-b). What is gene therapy?, 1–2.

Mandal, A. (1990). Gene Therapy History, 1–2. Retrieved from http://www.news-medical.net/health/Gene-Therapy-History.aspx

Migliorati, G., & Nicoletti, I. (1994). Pergamon 0145-2126(94)E0016-3, *18*(6), 431–439.

Nihongaki, Y., Yamamoto, S., Kawano, F., Suzuki, H., & Sato, M. (2015). CRISPR-Cas9-based Photoactivatable Transcription System. *Chemistry & Biology*, *22*(2), 169–174. http://doi.org/10.1016/j.chembiol.2014.12.011

Ran, F. A., Hsu, P. D., Lin, C. Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., … Zhang, F. (2013). Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity. *Cell*, *154*(6), 1380–1389. http://doi.org/10.1016/j.cell.2013.08.021

Robert, F., Barbeau, M., Éthier, S., Dostie, J., & Pelletier, J. (2015). Pharmacological inhibition of DNA-PK stimulates Cas9-mediated genome editing. *Genome Medicine*, 1–11. http://doi.org/10.1186/s13073-015-0215-6

Sally, R. (n.d.). Gene Therapy Types, 1–2.

Sander, J. D., & Joung, J. K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature Biotechnology*, *32*(4), 347–55. http://doi.org/10.1038/nbt.2842

Sun, N., & Zhao, H. (2013). Transcription activator-like effector nucleases (TALENs): A highly efficient and versatile tool for genome editing. *Biotechnology and Bioengineering*, *110*(7), 1811–1821. http://doi.org/10.1002/bit.24890

Vasileva, E., Shuvalov, O., Garabadgiu, A., Melino, G., & Barlev, N. (2015). Genome-editing tools for stem cell biology. *Cell Death and Disease*, *6*(7), e1831. http://doi.org/10.1038/cddis.2015.167

Verma, I. M., & Somia, N. (1997). Gene therapy – promises , problems and prospects, *389*(September), 239–242.

Werner, D. F., Blednov, Y. a, Ariwodola, O. J., Silberman, Y., Logan, E., Berry, R. B., … Homanics, G. E. (2006). Knockin Mice with Ethanol-Insensitive a1-Containing g -Aminobutyric Acid Type A Receptors Display Selective Alterations in Behavioral Responses to Ethanol. *The Journal of Pharmacology and Experimental Therapeutics*, *319*(1), 219–227. http://doi.org/10.1124/jpet.106.106161.the

Wu, S., Ying, G., Wu, Q., & Capecchi, M. R. (2008). A protocol for constructing gene targeting vectors: generating knockout mice for the cadherin family and beyond. *Nature Protocols*, *3*(6), 1056–1076. http://doi.org/10.1038/nprot.2008.70

Yuan, W., Leisner, T. M., McFadden, A. W., Wang, Z., Larson, M. K., Clark, S., … Parise, L. V. (2006). CIB1 is an endogenous inhibitor of agonist-induced integrin alphaIIbbeta3 activation. *The Journal of Cell Biology*, *172*(2), 169–75. http://doi.org/10.1083/jcb.200505131